The Bell System Technical Journal

Vol. XV *January, 1936* *No. 1*

Long-Wave Radio Transmission Phenomena Associated with a Cessation of the Sun's Rays

By AUSTIN BAILEY and A. E. HARPER

The variations in long-wave radio field strength near the time of sunset on long transmission paths, which have been reported by many observers, were studied for the purpose of formulating rational methods of forecasting their time of occurrence. During some preliminary observations fair agreement was found between the time of minimum field and the sun's position relative to a particular point on the transatlantic path under observation. The more extended study of radio field variations during sunset periods and solar eclipses disclosed that in general no exact relationship could be established between the sun's position at any point and the occurrence of the minimum field.

Observations of field variations were made on radio signals at a number of different frequencies and over several paths. It was concluded that a characteristic sunset cycle of field variations is present on frequencies between 18 kc. and 68 kc. for transmission paths longer than 700 km. For paths less than 200 km. long, such variations are negligible. There is some evidence that the amplitude of these field variations is smaller at lower frequencies.

Analysis of the data presented indicates that long waves over long paths are transmitted predominately by "sky waves." From the data it was not possible to establish any satisfactory picture of the path followed. It was established, however, that empirical methods based on observations over a particular transmission path may be employed to forecast the approximate time of occurrence of field variations.

Introduction

COMMERCIAL transatlantic radio telephone service to Europe was inaugurated on January 7, 1927, using a long-wave circuit on a frequency of 60 kilocycles. The difficulties of maintaining a satisfactory long-wave circuit through that part of each day when sunset conditions existed along the transmission path had been recognized [1] * for several years prior to the opening of service. After short-wave facilities were added to the transatlantic service, it became important to coordinate operating times of the short-wave transmitters in relation to the long-wave channel to assure maximum reliability and efficiency of service. In order to do this, a more precise knowledge of the

* Numbers refer to appended list of references.

behavior of the long-wave circuit during the sunset period was needed, and it was for this purpose that many of the observations reported in this paper were made.

An analysis of these observations indicated that, aside from their practical application, some rather fundamental information concerning the probable mechanism of transmission on long waves could be obtained. To these observations other data were added, and all of this available material was systematically studied to determine the effect of the cessation of the sun's active rays on radio transmission at long wave-lengths. Although the results alone are rather inconclusive, they do provide sufficient evidence to indicate that the mechanism of long-wave transmission is in some ways the same as that of short-wave transmission and for the longer paths depends primarily on waves returned to the earth by layers in the atmosphere.

No attempt will be made in this paper to review the present status of radio transmission theory or of related geophysical phenomena, except in special cases where required to show the significance of our results. A background of the theory has been provided by many investigators, among whom Smith,[2] Pedersen,[3] Anderson,[4] Appleton,[5] Green,[6] Namba,[7] Yokoyama and Tanimura,[8] Hollingworth [9] and Heising [10] should be mentioned.

The analysis of the data taken during the present investigation indicated that the connection between solar and radio phenomena is effected through the agency of electromagnetic radiation, and not by means of low velocity corpuscular solar emission. A partial corroboration of this conclusion was secured by Schafer and Goodall,[11, 12, 13] the U. S. Bureau of Standards,[14] and others during several solar eclipses. The rather complete analysis of data taken during eclipses which has been made by Appleton and Chapman [15] also confirms this view.

Method of Measurement and Estimated Precision

The field strength of special 60-kc. test dashes transmitted from WNL, Rocky Point, New York, was measured at Houlton, Maine, Chatham, New Jersey, and Cupar, Scotland. The Houlton and Chatham measurements were made by means of meter comparisons with a calibrated local oscillator and it is believed that their relative accuracy is of the order of $\pm$ 0.1 db, although the absolute accuracy probably falls short of this figure by a considerable margin. Due to the comparatively slow rate of long-wave field variation, no effort was exerted to obtain better timing than $\pm$ 10 seconds. Since weak fields and high noise are common during the late afternoon hours on the transatlantic path, it is believed that the precision of the Cupar measurements necessarily falls short of that possible for local tests.

Measurements made at Houlton on telegraph traffic from Tuckerton, Nauen, Ongar and Rocky Point are believed to be subject to relative errors as great as or even greater than ± 2.5 db. Comparisons between a local oscillator and telegraph traffic ordinarily are made by means of a cathode ray tube, and errors are occasioned both by the difficulty of reading the tube scale and by the varying strength of telegraph signals whose intensity is a function of the keying speed probably because of sluggish antenna systems.

The measurements on special 60-kc. test signals are believed to be sufficiently precise to provide a satisfactory index of the phenomena. The measurements on telegraph traffic provide data for a qualitative estimate of the nature of the various phases of the phenomena but probably are of no great value in fixing the exact time of occurrence of field strength increments smaller than 2 db, which may represent the total variation of some phases of the diurnal cycle.

General Characteristics of Sunset Effect on Short Paths

The form of the average diurnal sunset cycle of received field strength, plotted as a function of the sun's angular altitude at some salient point on the great circle transmission path, is shown on Fig. 1 for several paths. With the exception of the shortest path of 122 km. between Rocky Point and Chatham, New Jersey, a well defined typical sunset cycle was observed in all cases. If we assume that the sunset dip is due directly or indirectly to the cessation of the active solar rays in the upper atmosphere, the absence of characteristic large field variations at sunset for short paths may be ascribed to a predominance of low elevation transmission which does not enter the layers ionized by the sun. If rectilinear this transmission would pass ½ km. above Chatham on the 122-km. path, due to tangency with the earth's surface. The possibility of transmission between two points 122 km. apart on the earth's surface without a "sky wave," [5] therefore requires that the ray be bent around the spherical contour of the earth by diffraction and atmospheric refraction.[3, 16]

The daily occurrence of a sunset dip on paths of 700 km. and longer may be explained in several ways, all of which, however, require the existence of a downcoming ray. For example, the sunset phenomena may be due to a "fault" in the reflecting ionized layer,[7] or to interference between two reflected rays or between a reflected ray and a ground wave.[9, 17, 18]

Briefly, in accordance with the "fault" hypothesis, the shadow cast upon the ionized reflecting layer by the sun's active rays tangent to the earth's surface, or to an opaque atmospheric layer concentric with the earth which hereafter will be called the occulting layer, produces a

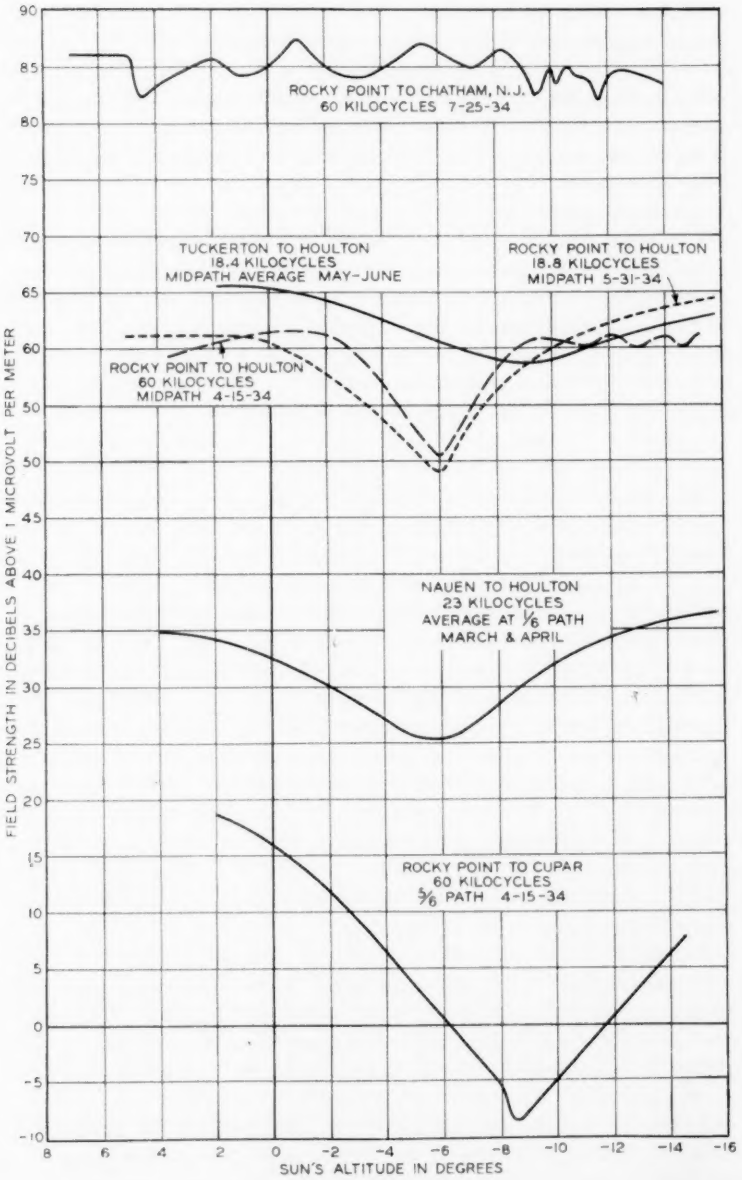


Fig. 1—Measured radio field strength plotted as a function of the sun's angular altitude at some salient path point.

"fault" in the otherwise uniform daylight ionized layer surface. The form of fault presumably is a ring generated by the intersection of the shadow cone of the occulting layer with the spherical reflecting layer surface. As the fault passes over the reflected radio ray apex, reception of the optimum ray may be impeded, either because of scattering due to an irregular reflecting surface,[3] or on account of the passage of the ionization density through a value defined by the Brewster angle, thereby changing the mode of transmission from daylight metallic reflection to night-time refraction.[7]

One of the objects of this analysis was to find the angular altitudes of the sun at some fixed point on the transmission path, coinciding with the phases of the sunset cycle of field variation at the receiving station. It was expected that with these data available over the period of a year, it would not be difficult to locate the salient point on the transmission path affected by the cessation of the sun's rays by means of Sumner lines of position [19, 20] provided, of course, that the physical structure of the layers showed negligible variation. From the above-mentioned data it would also be possible to compute the distance between the occulting and reflecting layers [21] and to accurately predict the time of occurrence of the sunset minimum. As shown in Tables I–IV, annual and fortuitous variations in the time of occurrence of the phases of the phenomenon prevent the satisfactory application of this method to long path effects, and when applied to the Rocky Point-Houlton path the point so located apparently is situated a considerable distance to the southwest of the most southerly path terminal. This presumably indicates that, if the phenomena take place at a fixed location on the transmission path, there must either be an annual variation of considerable magnitude in the effective distance between the reflecting and occulting layers, or the mechanism involved must be considerably more complex than that initially assumed.

Frequency Range of Phenomena

The characteristic diurnal sunset cycle was found on all frequencies studied during this investigation, the scope of which was limited to frequencies between 18 and 68 kilocycles. Available published data taken by other investigators indicate that the pronounced sunset minimum characteristic of long-wave transmission is not observed at broadcast frequencies.[6, 22] This may be due to the failure of transmission to improve after the sunrise drop, giving an all day minimum, or, if the minimum is due principally to interference bands, the fineness of band pattern at high frequencies may prevent the phenomenon from being recognized.

Observations on GBR at 16 kilocycles at Houlton during the latter part of April, 1934, showed little evidence of a minimum. These measurements were made on telegraph traffic and if a minimum occurred its amplitude must have been less than the observational errors of this method of measurement. These results seem to be confirmed by the observations of Yokoyama and Tanimura [8] which show a pronounced decrease in the amplitude of the sunset cycle at frequencies below 17 kilocycles, while frequencies slightly above this value display the characteristic effects. This apparent difference in the behavior of 18 and 16-kc. transmission, seems rather unusual and if real may have some geophysical significance.

Results

An examination of the data taken on the Rocky Point-Houlton 60-kc. transmission path as plotted on Figs. 2 and 3 discloses that of the ten cases plotted, nine show that a minimum in measured field occurs 22–26 minutes or an average of 23 minutes after surface sunset at Rocky Point, corresponding to an altitude of the sun of 4 to 5½ degrees below the horizon, and in a single case at 18 minutes after sunset on August 29, 1934. The results in the winter months seem less regular than in summer and the cases of 2/25/34 and 1/21/34 are especially noteworthy in that they show multiple minima, the 23-minute minimum being subsidiary to a minimum occurring about an hour after Rocky Point sunset on 1/21/34. The fact that minima occur at a nearly constant interval after Rocky Point surface sunset, which occurs at the same time that the sun's rays are tangent to concentric elevated layers above this point, rather than sunset at some other path point, is believed to be the result of a fortuitous combination of circumstances rather than a rational relationship between these times. This hypothesis is strengthened by data shown in Table I taken over the same path at 18 kilocycles and data in Table II taken at 18 kilocycles over the 900-km. Tuckerton-Houlton path which do not show this same constant relationship with the time of sunset at the transmitting station. Of three observations obtained over the Rocky Point-Houlton 18-kc. transmission path the minimum in field occurred 34 minutes after sunset at Rocky Point in two cases but on the other day it occurred 22 minutes after sunset at Rocky Point. On all three occasions, however, the minimum was between 35 and 37 minutes ($-6°$ to $-6.5°$ altitude of sun) after surface sunset at the mid-path point. As in the case of the 60-kc. path for which data are shown on Figs. 2 and 3, the field began to fall at mid-path surface sunset.

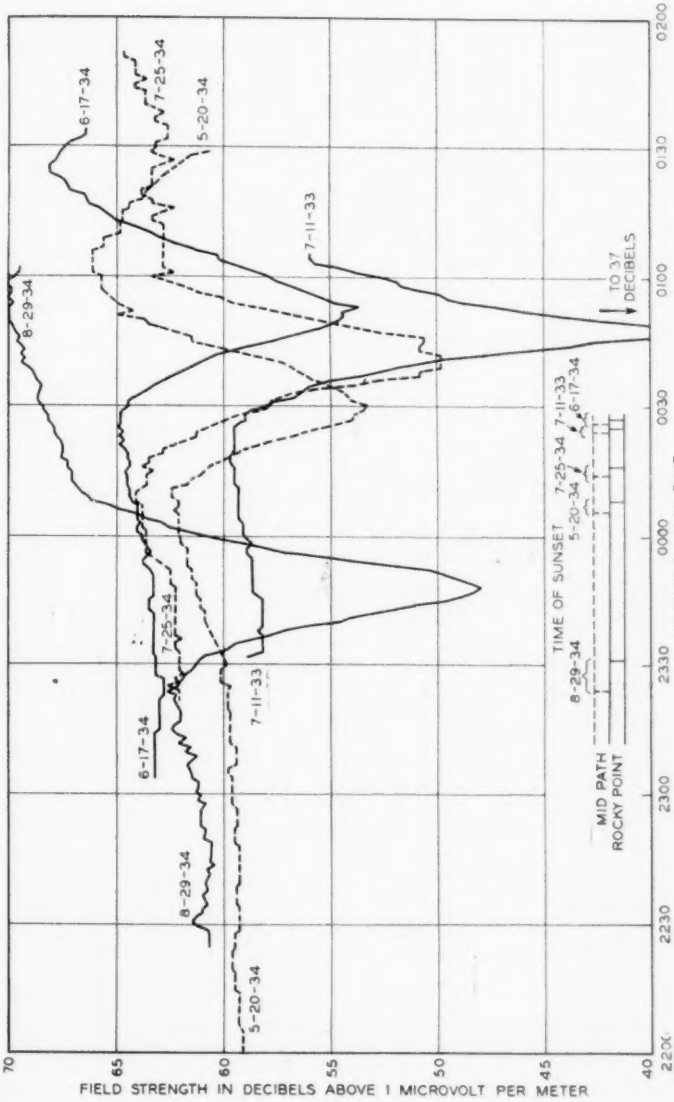


Fig. 2—Field strength variations on the Rocky Point-Houlton 60 kc. transmission path during the summer months.

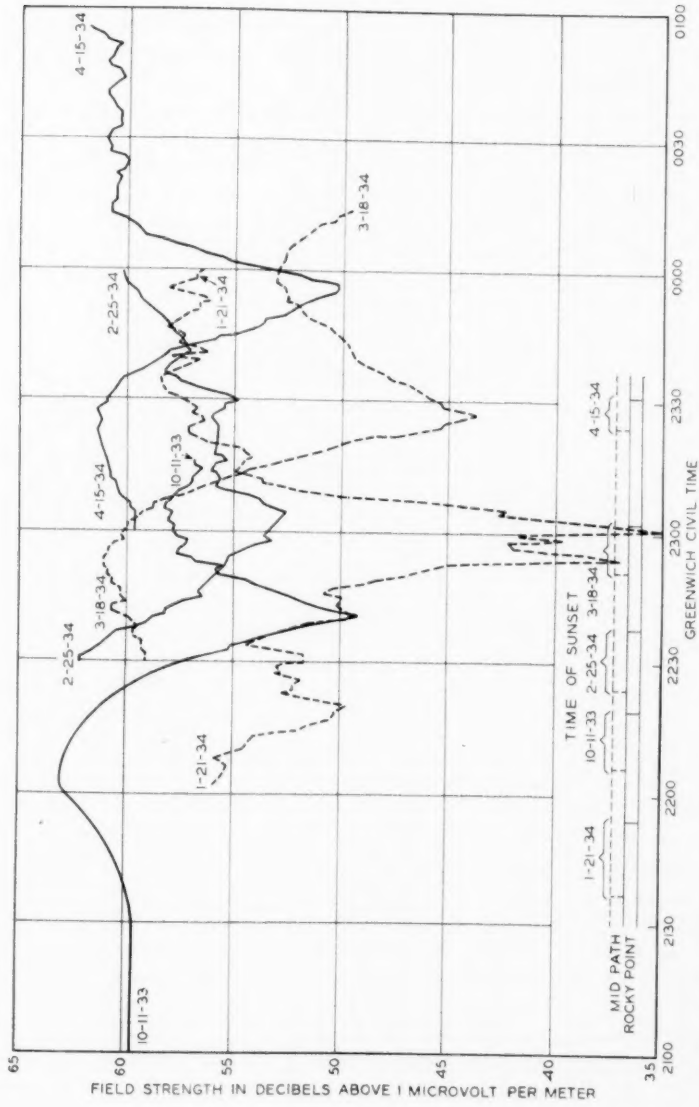


Fig. 3—Field strength variations on the Rocky Point-Houlton 60 kc. transmission path during the winter months.

TABLE I

ROCKY POINT-HOULTON—18.8 KC.

Distance 700 Km. Average Minimum 13 Db Below Field Immediately Preceding Drop

Date	Time of Sunset, GCT		Minimum Field		Interval Between Sunset and Minimum		
	Rocky Point	Houlton	Time GCT	Sun's Altitude at Midpath Degrees	Midpath	Rocky Point	Houlton
6/ 1/34	0016	0014	0050	−6.0	0:35	0:34	0:36
8/ 2/34	0009	0003	0043	−6.7	0:37	0:34	0:40
10/21/34	2204	2137	2226	−7.4	0:35	0:22	0:49
			Average	−6.7	0:36	0:30	0:42

TABLE II

TUCKERTON-HOULTON—18.4 KC.

Distance 900 Km. Average Minimum 7.9 Db Below Field Immediately Preceding Drop

Date	Time of Sunset, GCT		Minimum Field		Interval Between Sunset and Minimum		
	Tuckerton	Houlton	Time GCT	Sun's Altitude at Midpath Degrees	Midpath	Tuckerton	Houlton
4/28–29/34	2349	2335	0030	− 9.0	0:48	0:41	0:55
4/30–5/1/34	2351	2337	0033	− 9.0	0:49	0:42	0:56
5/ 1– 2/34	2352	2338	0040	−10.0	0:55	0:48	1:02
5/ 3– 4/34	2354	2342	0045	− 9.6	0:57	0:51	1:03
5/ 9–10/34	2359	2349	0053	−10.2	1:00	0:54	1:04
5/10–11/34	0000	2350	0047	− 9.4	0:52	0:47	0:57
5/11–12/34	0001	2351	0027	− 5.9	0:31	0:26	0:36
5/17–18/34	0006	2359	0048	− 8.0	0:46	0:42	0:49
5/21–22/34	0010	0004	0046	− 6.9	0:39	0:36	0:42
5/24–25/34	0013	0007	0050	− 7.0	0:40	0:37	0:43
5/28–29/34	0016	0011	0115	− 9.8	1:02	0:59	1:04
6/ 4– 5/34	0021	0017	0117	− 9.1	0:58	0:56	1:00
6/ 7– 8/34	0023	0019	0115	−11.0	0:54	0:52	0:56
6/11–12/34	0025	0022	0120	− 9.0	0:57	0:55	0:58
6/14–15/34	0027	0023	0122	− 9.3	0:57	0:55	0:59
			Average	− 9.5	0:51	0:47	0:55

The 18-kc. Tuckerton-Houlton 900-km. path of nearly the same azimuth as the Rocky Point-Houlton path, showed a minimum occurring irregularly at from 31 to 62 minutes, or an average of 51 minutes after mid-path sunset, corresponding to an average sun's altitude of − 9.5°. Since the data of Table II are subject to the errors inherent

to measurements made on telegraph traffic, a portion of the large variation in time may be due to experimental errors.

The wave-like variations in field intensity observed on the Rocky Point-Houlton 60-kc. path on 1/21/34 and 4/15/34 have the appearance of interference fringes,[9, 17] and if explained on this basis the question at once arises whether or not interference is the principal cause of the sunset cycle. If daylight communication is accomplished either by means of a single reflected ray in combination with a ground wave, or by the resultant of a number of reflected rays of nearly constant complex propagation constants, at sunset the occultation of the ionizing rays from the transmission medium by tangency with an occulting layer, might reasonably be expected to produce variations in both the real and imaginary portions of the propagation constant of the medium, thereby causing interference fringes through the variation of the relative phase of different rays.

As an example, we may assume that on the Rocky Point-Houlton transmission path the ground ray provides the principal agency of daylight communication by means of atmospheric refraction and diffraction. At the approach of sunset, reduced ionization between the earth and the reflecting layer reduces the attenuation in the path of the reflected ray, producing a resultant received field which is a function of the relative intensity and phase of the two waves. As the effect of the sun's active rays becomes still less, the decreased ionization of the layers produces variations in the phase of arrival of the reflected wave, either through changes in the propagation constant, or on account of greater path length occasioned by an increased virtual height of the reflecting layer.

Now since the resultant received field is the vector sum of the two rays, when one ray is much smaller than the other, variations in their relative phase will produce small amplitude fluctuations in the resultant, with maxima and minima equal to the sum and difference of the two components. As the two components approach equality, however, the maxima will approach twice the intensity of one component ray, while the minima will approach zero, thereby producing a very deep "dip" in the received field.

The "D" Layer

It has been suggested by Heising,[10] Appleton [23] and Chapman [21] that passage through a low-elevation layer of ozone produced by solar ionization, is one of the principal causes for the daylight attenuation of the reflected ray.[2] Radio transmission measurements at broadcast frequencies, reported by the U. S. Bureau of Standards [22] and by the Australian Council for Scientific and Industrial Research,[6] show a rapid

decrease in sky-wave attenuation beginning shortly before tangential sunset and continuing for from one-half to two hours after sunset. On the 60-kc. long-distance path the evidence of the initial phases of this phenomenon is not so well defined, due probably to the gradual nature of the change, and to the characteristic sunset minimum which occurs shortly after sunset. Recent measurements made on the transatlantic radio telephone channels, however, indicate a presunset rise in field of about 2 db, beginning when the sun's altitude becomes low enough to cause an appreciable lengthening of the atmospheric ray path, thereby decreasing the intensity of the active solar rays. A presunset increase is very apparent on the Rocky Point-Houlton 60-kc. measurements, beginning at about 40 minutes before, and continuing to a maximum at the instant of mid-path sunset. There is a possibility, however, that this phase may be due to interference phenomena.

More conclusive evidence of the effect of a reduction in the intensity of the sun's ionizing rays is provided by long-wave field strength measurements made during the solar eclipses of January 24, 1925, and August 31, 1932. The data taken in 1925 were secured by means of an automatic recorder on the Rocky Point-Belfast 57-kc. path and manual measurements were made at the European receiving stations. In 1932, automatic recorders were used on both the Rocky Point-Houlton 60-kc. path and the Rugby-Houlton 68-kc. path. In all cases where automatic recorders were used the data were abstracted from the record and replotted for reproduction.

Figures 4 and 5 show the eclipse circumstances and the concurrent variations in the measured radio field strength. The Chedzoy, New Southgate, and somewhat less clearly the Aberdeen measurements in 1925, show the results of reduced attenuation almost immediately after the first darkening of the transmission path. For the 1932 data this effect is clearly evident on the Rocky Point-Houlton measurements.

The sudden increase in field to be expected at the beginning of the eclipse at Belfast in 1925 is not apparent from the data, and its absence may be due either to improper recorder operation or to some fortuitous phenomenon peculiar to that particular eclipse, such as, for example, the possibility that the phases of ground and sky waves were in quadrature. The Rugby-Houlton observations in 1932 likewise are ambiguous because the true eclipse effect is complicated by superposed sunset effects originating at the eastern path terminal, and the observed increase in field may be due to these rather than to the eclipse.

In all cases except the two mentioned above, the increase in field was followed by a rapid drop, with a minimum occurring at the approximate time the totality shadow crossed the transmission path.

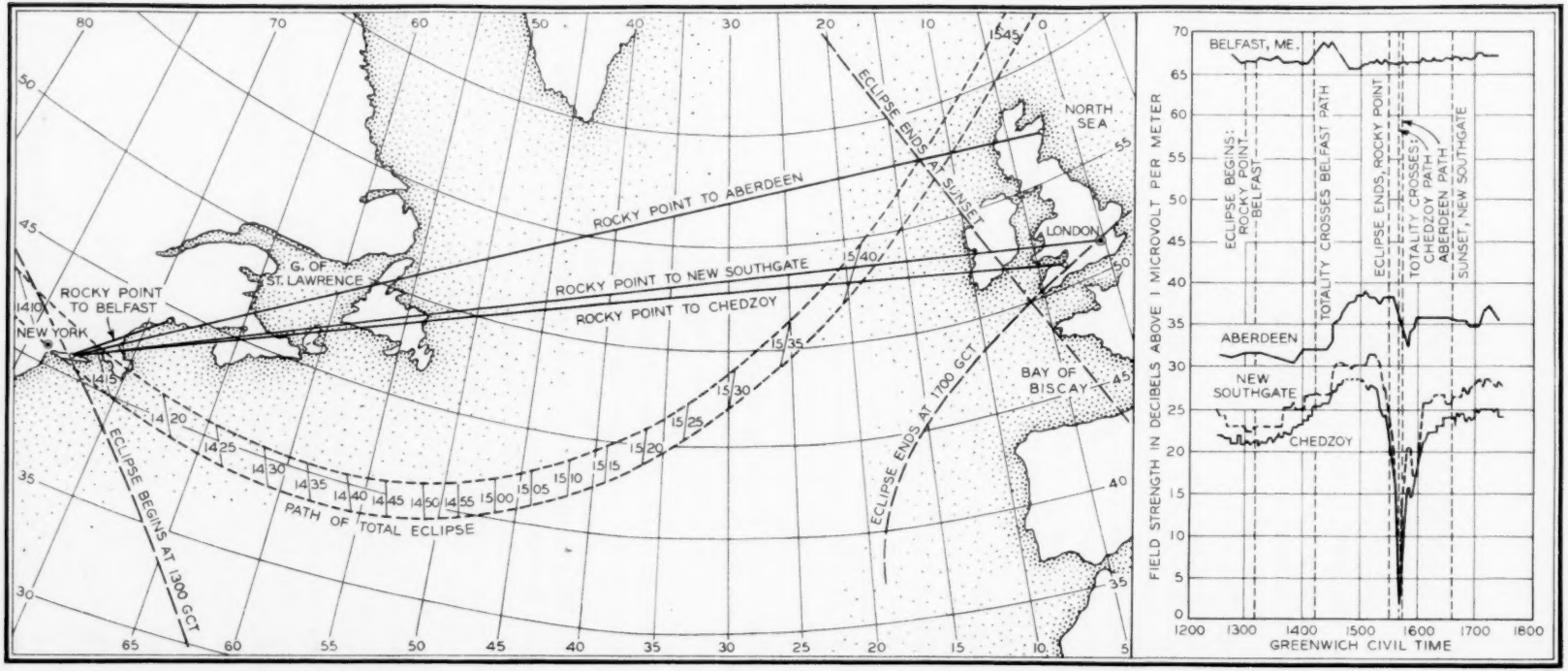


Fig. 4—Relation between 57 kc. radio transmission and circumstances of the solar eclipse of January 25, 1925.

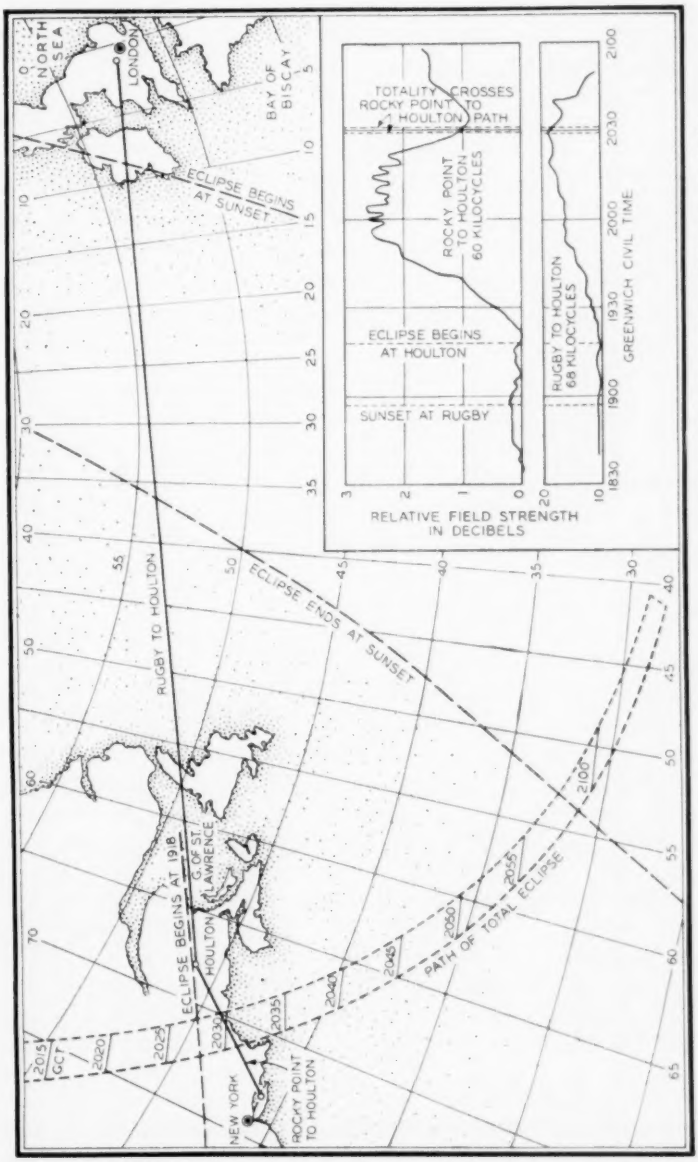
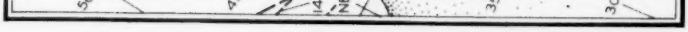


Fig. 5—Relation between 60 kc. radio transmission and circumstances of solar eclipse of August 31, 1932.



Long Transmission Paths

An examination of long-distance data of Tables III and IV discloses the following facts regarding these phenomena.

TABLE III

Nauen-Houlton—23 Kc.

Distance 5600 Km. Average Minimum 13 Db Below Field Immediately Preceding Drop

Date	Time of Sunset, GCT		Minimum Field		Interval Between Sunset and Mininum	
	Nauen	1/6 Path Point	Time GCT	Sun's Altitude at 1/6 Path Point Degree	Nauen	1/6 Path Point
3/12/34.....	1706	1756	1838	−6.2	1:32	0:42
3/19/34.....	1718	1810	1836	−4.2	1:18	0:26
3/22/34.....	1723	1816	1900	−5.5	1:37	0:44
3/26/34.....	1730	1824	1850	−4.2	1:20	0:26
3/29/34.....	1735	1830	1903	−5.5	1:28	0:33
4/ 2/34.....	1742	1839	1925	−6.7	1:43	0:46
4/ 5/34.....	1747	1845	1857	−2.5	1:10	0:12
4/ 9/34.....	1754	1853	1940	−7.0	1:46	0:47
4/10/34.....	1756	1855	1944	−7.0	1:48	0:49
4/12/34.....	1759	1859	1950	−7.2	1:51	0:51
			Average	−5.6	1:33	0:37

TABLE IV

Rocky Point-Cupar—60 Kc.

Distance 5172 Km. Average Minimum 25 Db Below Field Immediately Preceding Drop

Date	Time of Sunset, GCT		Minimum Field		Interval Between Sunset and Minimum	
	Cupar	5/6 Path Point	Time GCT	Sun's Altitude at 5/6 Path Point Degrees	Cupar	5/6 Path Point
10/11/33....	1724	1817	1910	− 8.0	1:46	0:53
11/22/33....	1554	1644	1807	−10.4	2:13	1:23
12/20/33....	1548	1626	1756	−10.6	2:08	1:30
2/25/34....	1736	1830	1958	−12.5	2:22	1:28
3/18/34....	1820	1916	2029	−10.7	2:09	1:13
4/15/34....	1917	2015	2117	− 8.4	2:00	1:02
6/17/34....	2101	2206	2322	− 6.6	2:21	1:16
			Average	− 9.6	2:08	1:15

Rocky Point-Cupar 60-Kc. Path, Table IV

1. Minimum field does not follow surface sunset at any point on the transmission path by a constant time interval independent of season.

2. The altitude of the sun as measured from the most easterly path apex of an hypothetical three-reflection path varies from approximately $-6°$ in summer to $-13°$ in winter at the instant of minimum field, and computation shows that there is no point on the path at which the altitude remains constant.

3. Daily irregularities of considerable magnitude are often noted.

Ongar-Houlton Path

1. Multiple minima of irregular distribution seem characteristic of this path.

2. Averages of field strength as a function of the sun's altitude at three path apices show separate minima when the sun is at about $-6°$ during May and June.

Nauen-Houlton, Table III

1. The time of minimum field varies as much as 40 minutes within a few days.

2. An average of the March-April data indicates that the minimum takes place when the sun's altitude is approximately $-5.6°$ at the first of three path apices, or 37 minutes after sunset at the earth's surface under this point.

Conclusions

Based solely upon the rather meagre data gathered during this study, and therefore subject to further confirmation before being considered generally applicable to all long-wave transmission paths, our conclusions may be recapitulated as follows:

1. A characteristic cycle of events accompanies the cessation of the sun's active rays, consisting of an initial increase in field during the reduction of solar intensity, followed by a minimum received field after the sun's rays are cut off at the earth's surface. This cycle, occurring both at sunset and during total solar eclipses, is typical of long-wave transmission in the 18–70-kc. band over paths in excess of a few hundred kilometers.

2. The time interval between sunset at any point on the transmission path and the instant of minimum field has an annual and an apparently fortuitous daily variation. As a result of these variations the minimum does not occur when the sun is at a fixed angular altitude at any point on the transmission path.

3. The time interval between sunset at some fractional path point and the instant of minimum field, and likewise the angular altitude of the sun referred to the plane of the horizon at such a point, increases with both the path length and the wave-length, and is maximum during the winter months.

4. The amplitude of variation of the sunset cycle apparently is reduced greatly at frequencies below 17 kilocycles and on paths shorter than 200 km.

5. Evidence of interference fringes on some of the observations suggests the possibility that the main sunset minimum may be the result of interference phenomena.

6. The fact that the phases of the radio transmission cycle closely follow the optical eclipse circumstances indicates that the radio phenomena must be related to solar radiation of velocity similar to that of light.

7. On transatlantic paths during the spring and summer months the average sunset minimum occurred when the sun was approximately 6° below the horizon at one-sixth of the total path length from the eastern terminal. Since these paths varied considerably in latitude and length, the phenomena may be related to effects occurring at the most easterly apex of a three-reflection path. The data obtained in this investigation, however, are not sufficient to definitely establish the generality of this hypothesis.

8. Empirical rules may be formulated for the prediction of the time of occurrence of various phases of the sunset cycle on short transmission paths. For example, the beginning of the drop in field on the 60-kc. Rocky Point-Houlton path occurs at mid-path surface sunset, and the minimum occurs at approximately 23 minutes after surface sunset at Rocky Point. On longer paths larger fortuitous variations occur and available data fail to connect the time of the minimum with sunset at any point on the transmission path. Representative curves drawn from the data, although subject to random errors, provide an empirical method for the prediction of the approximate time of occurrence of the phenomena and are of service in traffic and power scheduling.

Acknowledgments

The authors are indebted to the engineers of the British Post Office and the operating personnel of the American Telephone and Telegraph Company, for their generous cooperation in measurement schedules, and to their colleagues in Bell Telephone Laboratories for their assistance and constructive criticism in the preparation of this paper.

References

1. "Transatlantic Radio Telephony," H. D. Arnold and Lloyd Espenschied, *Jour. A.I.E.E.*, August 1923; *Bell Sys. Tech. Jour.*, October 1923.
2. 24th Kelvin Lecture, "The Travel of Wireless Waves," by Sir Frank E. Smith, *Inst. E.E.*, Vol. 9, No. 25, March 1934.
3. "The Propagation of Radio Waves," by P. O. Pedersen.
4. "Transatlantic Radio Transmission and Solar Activity," C. N. Anderson, *Proc. I.R.E.*, Vol. 16, pp. 297–347, March 1928.

5. "On Some Direct Evidence for Downward Atmospheric Reflection of Electric Waves," by E. V. Appleton and M. A. F. Barnett, *Proc. Royal Soc.*, 109, 1925, pp. 621–641.
6. "A Preliminary Investigation of Fading in New South Wales," by A. L. Green and W. G. Baker, Bulletin No. 63, Radio Research Board Report No. 4, Council for Scientific and Industrial Research, Commonwealth of Australia, 1932.
7. "General Theory of the Propagation of Radio Waves in the Ionized Layer of the Upper Atmosphere," Shogo Namba, *Proc. I.R.E.*, February 1933.
8. "Some Long Distance Transmission Phenomena of Low-Frequency Waves," I. Yokoyama and I. Tanimura, *Proc. I.R.E.*, February 1933.
9. "The Propagation of Radio Waves," by J. Hollingworth, *Jour. I.E.E.*, Vol. 64, 1925–1926, pp. 579–589.
10. "Experiments and Observations Concerning the Ionized Regions of the Atmosphere," by R. A. Heising, *Proc. I.R.E.*, Vol. 16, January 1928, pp. 75–99.
11. "The Effect of the Recent Solar Eclipse on the Ionized Layers of the Upper Atmosphere," Schafer and Goodall, *Science*, Vol. 76, November 11, 1932, pp. 444–446.
12. "Radio Studies of the Ionosphere," by Schafer and Goodall, *Nature*, September 30, 1933, pp. 521–522.
13. "Ionosphere Measurements During the Partial Eclipse of the Sun of February 3, 1935," by Schafer and Goodall, *Nature*, March 9, 1935, pp. 393–394.
14. "Radio Observations of the Bureau of Standards During the Eclipse of August 31, 1932," by S. S. Kirby, L. V. Berkner, T. R. Gilliland and K. A. Norton, *Proc. I.R.E.*, Vol. 22, No. 2, February 1934.
15. "Report on Ionization Changes During a Solar Eclipse," by E. V. Appleton and S. Chapman, *Proc. I.R.E.*, Vol. 23, June 1935, pp. 658–669.
16. "Radio Propagation Over Spherical Earth," by C. R. Burrows, *Proc. I.R.E.*, Vol. 23, May 1935, pp. 470–480.
17. "Summary of Progress in the Study of Radio Wave Propagation Phenomena," by G. W. Kenrick and G. W. Pickard, *Proc. I.R.E.*, Vol. 18, April 1930, pp. 649–668.
18. "Phase Interference Phenomena in Low-Frequency Radio Transmission," by G. W. Kenrick and G. W. Pickard, *Proc. I.R.E.*, Vol. 22, March 1934, pp. 344–358.
19. "American Practical Navigator," Bowditch, *Hydrographic Office Publication H.O.* No. 9, Government Printing Office, Washington, D. C.
20. "The Sumner Line of Position," *U. S. Hydrographic Office Publication H.O.* No. 203, Government Printing Office, Washington, D. C.
21. Bakerian Lecture, "Some Phenomena of the Upper Atmosphere," by S. Chapman, *Proc. Royal Soc.*, Vol. 132, No. A820, August 1, 1931.
22. "An Analysis of Continuous Records of Field Intensity at Broadcast Frequencies," by K. A. Norton, S. S. Kirby and G. H. Lesler, Research Paper RP752, *Jour. of Research, Nat. Bureau of Standards*, Vol. 13, December 1934.
23. "Some Measurements of the Equivalent Height of the Atmospheric Layer," by E. V. Appleton, *Proc. Royal Soc.*, Vol. 126, 1929–1930, p. 543.

APPENDIX I

Position of Intermediate Points on the Transmission Path

The method of determining the transmission path parameters and the position of intermediate path points spaced a given distance from a path terminal is given in detail below.

Let φ = Latitude,
L_o = Longitude,
D = Distance, nautical miles between subscript points,
C = Path azimuth at subscript point,

Subscript a denotes sending station,

Subscript b denotes receiving station,
Subscript x denotes intermediate point,
L_{oab} = Difference in longitude between a and b.

By the law of cosines, in a spherical triangle of sides abc and angles ABC, if we are given two sides and the included angle, the side opposite the given angle may be computed from (1) below.

$$\cos a = \cos b \cos c + \sin b \sin c \cos A \tag{1}$$

or substituting geographical coordinates

$$\cos D_{ab} = \sin \varphi_a \sin \varphi_b + \cos \varphi_a \cos \varphi_b \cos L_{oab}. \tag{2}$$

This may be made more convenient for logarithmic computation by writing it in the following form:

$$\text{hav } D_{ab} = \text{hav } (\varphi_a - \varphi_b) + \cos \varphi_a \cos \varphi_b \text{ hav } L_{oab}. \tag{3}$$

Now by the law of sines

$$\begin{cases} \sin C_a = \dfrac{\cos \varphi_b \sin L_{oab}}{\sin D_{ab}} = \cos \varphi_b \sin L_{oab} \csc D_{ab}, \\ \sin C_b = \dfrac{\cos \varphi_a \sin L_{oab}}{\sin D_{ab}} = \cos \varphi_a \sin L_{oab} \csc D_{ab}. \end{cases} \tag{4}$$

Equations (3) and (4) above determine the great circle distance between "a" and "b," the azimuth of "a" from "b," and "b" from "a." To find the position of a point "x" located a fraction of the total distance between "a" and "b" we again substitute in (1), obtaining

$$\begin{cases} \sin \varphi_x = \sin \varphi_b \cos D_{xb} + \cos \varphi_b \sin D_{xb} \cos C_b, \\ \sin \varphi_x = \sin \varphi_a \cos D_{xa} + \cos \varphi_a \sin D_{xa} \cos C_a, \end{cases} \tag{5}$$

and by the law of sines

$$\begin{cases} \sin L_{oax} = \dfrac{\sin C_a \sin D_{xa}}{\cos \varphi_x} = \sin C_a \sin D_{xa} \sec \varphi_x, \\ \sin L_{obx} = \dfrac{\sin C_b \sin D_{xb}}{\cos \varphi_x} = \sin C_b \sin D_{xb} \sec \varphi_x. \end{cases} \tag{6}$$

The latitude and longitude of intermediate point "x" are therefore determined by equations (5) and (6) above.

APPENDIX II

Determination of Sun's Altitude and Azimuth

The angle of the sun to the horizon and to the meridian plane at any hour may be computed by methods similar to the above. For this case we have a celestial triangle whose sides are a meridian through the observer's zenith, a meridian through the sun, and a great circle through the sun and the zenith. The arc subtended by the pole and zenith is the complement of the observer's latitude "φ," the arc subtended by the pole and the sun is the complement of the sun's declination "d" (celestial latitude), and the angle "t" at the pole between these two arcs is the sun's hour angle. With these two sides and included angle we may compute the arc between the sun and the zenith (complement of the altitude "h") and the sun's azimuth "z" which is the angle between the meridian containing the zenith and the great circle passing through the zenith and the sun.

By the law of cosines (1) above

$$\sin h = \sin d \sin \varphi + \cos d \cos \varphi \cos t \tag{8}$$

and by the law of sines

$$\sin Z = \frac{\sin t \cos d}{\cos h}. \tag{9}$$

Values of h and z as a function of φ, d and t are tabulated in convenient form in *hydrographic office publication H.O.* No. 203. The sun's declination and the computed times of sunset may be obtained from the *American Nautical Almanac*.

The Corrosion of Metals—I. Mechanism of Corrosion Processes

By R. M. BURNS

This paper outlines the application of electrochemical methods to corrosion investigations. It discusses the position of the potential of a metal against its environment and the trend of this potential with time, pointing out that it is thereby possible to determine whether the corrosion process is controlled by reactions occurring at the anodic areas, the cathodic areas, or both; that is, whether there is a tendency toward passivity, inhibition or progressive attack. Measurements of film stability whether in terms of the leakage current which may be passed through the film or in terms of the amount of film forming material required to produce passivity or the amount of film destroying material required to render a metal active, furnish information as to the quality of corrosion resistant films. Measurements of the rate at which a film forms on a metal when placed in a film-forming environment throws light on its relative surface reactivity, and such information is of assistance in determining the rate of corrosion in homogeneous corrosive environments or the rate of passivation in the film-forming environments. On the basis of such measurements and with a chemical knowledge of the environments in which metals are used as well as the composition and physical state or structure of the metals, it is possible to predict corrosion behavior and to obtain an understanding of corrosion problems usually not possible by ordinary empirical corrosion tests.

ALL metals are corrodible under the appropriate circumstances. The most important metal industrially, iron, is probably the most corrodible under ordinary conditions. Many estimates have been made of the value of iron and steel products destroyed by corrosion.[1] While much depends upon the basis of calculation it seems reasonable to conclude that the annual cost of corrosion in this country is of the same order as the interest on the public debt or nearly one third of the cost of the federal government in normal times. The common non-ferrous metals—zinc, lead, copper, aluminum, nickel and tin—are more resistant to corrosion largely because of their tendencies to form protective surface films. In the atmosphere under favorable circumstances tests have indicated, for example, that in the form of sheet 0.03 inch in thickness and exposed on one side as in the case of roofings, zinc, copper and lead if mechanically undisturbed would resist corrosion for more than one, two and three centuries respectively.[2] Once a protective film is formed it may preserve the metal indefinitely. Under other circumstances these metals may readily corrode. Contact with large inert soil particles may result in the perforation of cable sheathing 0.10 inch in thickness in about 8 years.[3] Tin, although resistant to corrosion in air and pure water, is severely corroded by alkalies, and aluminum is attacked by both alkalies and acids. The

noble metals such as gold, silver and platinum, being less reactive chemically than the more basic metals, are as a group the least corrodible, yet silver tarnishes markedly in moist atmospheres containing volatile sulfur compounds; gold is attacked by halogens in solution and platinum by fused alkalies.

The protection of metals from corrosion may be accomplished in general either by maintaining a non-corrosive surrounding environment or by coating the metallic parts with paints, lacquers or more corrosion-resistant metals. Such measures as the control of humidity and dust in interior atmospheres, deoxygenation of boiler waters, the use of passivators such as chromates, carbonates, phosphates, silicates and alkalies in the water-cooling systems or water scrubbers of air conditioning equipment and the use of cathodic protection which consists in setting up an electrolytic cell in which the metallic part subject to corrosion is made the cathode, are typical examples of environmental control designed to inhibit corrosion. Another well-known example of avoiding corrosion by control of environment is the protection of underground cables which is afforded by the proper drainage of electrical stray currents which have been picked up by the cable network. Where it is infeasible to maintain an inert environment the use of nonferrous metallic coatings is of great value in the preservation of steel products. Much metallurgical work has been devoted in recent years to the development of corrosion-resistant alloys. In both of these cases the protective feature consists in a naturally developed surface film. Where natural films afford insufficient protection it becomes necessary to resort to coatings of organic materials such as paints, lacquers, enamels, complex structures of such materials as pitches or asphalts with jute felt or paper, etc. It has been estimated that one hundred and twenty million gallons of paint are used annually for corrosion prevention.[4]

Corrosion may be defined in most general terms as the chemical reaction of a metal with the non-metallic constituents of its environment. In this sense any reaction in which a metal is degraded to one of its compounds, such as an oxide, hydroxide, acid or salt is a corrosion reaction. The nature of the reaction which occurs in any given case depends both upon the reactivity of the metal, its purity, physical state and surface condition and upon the character of the environment, that is, upon the chemical components present, their physical phases and concentrations. It also depends upon the temperature. Corrodibility is not wholly an inherent property of a metal which can be determined by a single arbitrarily chosen corrosion test of any sort; even the relative order of corrodibility of a series of metals

is not constant.[5] For example, iron, magnesium and zinc corrode in conductivity water exposed to the air in the order given; in sodium chloride solution the order is magnesium, zinc and iron, while finally in strong alkali solutions these metals corrode in the order: zinc, iron and magnesium.

It is apparent that the occurrence of corrosion depends upon both the character of the environment and of the metal. While the environment in which a metal is used is usually complex, it is generally possible to recognize those constituents which exert a controlling influence on the course of corrosion. In the ordinary atmosphere water vapor and oxygen are major factors in the process. Other substances such as sulfur dioxide, chloride ions and dust also influence the character of the corrosive attack. The green patina, a basic sulfate, which forms in the course of time on copper exposed to the air and which may preserve the metal for centuries, owes its origin to traces of sulfur dioxide in the atmosphere.[6] On the other hand, the higher concentrations of sulfur dioxide prevailing in the neighborhood of smelters which treat sulfur bearing ores may rapidly corrode copper telephone wires to destruction. Steel containing up to 0.25 per cent copper is about two-fold more resistant to corrosion than non-copper bearing steel in most industrial atmospheres; in New York City, however, the very small chloride ion content of the atmosphere, an otherwise typical industrial atmosphere, largely destroys the corrosion resistance conferred by the copper. Rainfall is an influential factor in the corrosion of metals exposed to the atmosphere. It may increase corrosion by removing soluble corrosion products from the surface of the metal, or it may retard corrosion by washing away dust particles and electrolytes, both of which promote corrosive attack. For example, in New York City the daily application of a water spray increased the rate of corrosion of zinc by 30 per cent but decreased the rate of corrosion of different ferrous materials from 30 per cent to 46 per cent, the amount of corrosion being determined by loss of weight measurements. The corrosion products of zinc are appreciably more soluble than those of iron and presumably were largely removed by the frequent washing. On the other hand, the deliquescent nature of the corrosion products of iron [7] at humidities prevailing a large part of the year provide, in contrast to the relatively less deliquescent corrosion products of zinc, a film of moisture more or less saturated with corrosive electrolytes and dust particles. This film is diluted or otherwise removed by the water spray.

Indoor environments differ from the outside atmosphere mainly in being drier, cleaner and subject to less pronounced temperature varia-

tions. As a consequence, metals corrode somewhat less rapidly indoors but the character of the attack generally resembles that shown out of doors. Underground exposures to soils and waters are often severe and cause ferrous structures to fail unless protective nonmetallic coatings are employed. Considerable progress has been made within the past ten years in determining the corrosivity of soils and developing adequate preservative coatings.[8]

The best measure of the tendency of a metal to corrode is, in thermodynamical terms, the decrease in free energy which accompanies the chemical reaction involved in the process, i.e., the difference in energy between the initial and final state of the system. This may be obtained by simple calculation and is of value in showing whether or not it is possible for corrosion to occur under the conditions defined. There is no assurance however, that reactions which are possible will actually take place within a reasonable time, if at all. Calculations [9] show that, exposed to the atmosphere containing moisture, aluminum, zinc, tin, iron, nickel, copper and silver may corrode to their respective hydroxides. If, oxygen be excluded the last three metals listed cannot corrode and iron only to the lower state of oxidation. These calculations, however, give no information as to the rate of corrosion or the mechanism by which it takes place, matters of great practical importance.

The rate of corrosion or of any other chemical reaction bears no direct relationship to the energy changes involved; it cannot be predicted but must be measured in some form.[10] Obviously the rate of corrosion depends upon the nature of the chemical reactions at the surface of the metal. Generally, secondary reactions are involved and the slowest step in the process controls the rate. The limiting factor is usually some sort of barrier,—a film of gaseous or solid corrosion products at the surface.

The mechanism by which corrosion occurs may be one either of direct combination of the metal and non-metal or the replacement by the corroding metal of hydrogen or another metal in compounds. The oxidation of metals, particularly at higher temperatures, halogenation reactions, such as the chlorination of aluminum, and the reaction of copper and sulfur, are examples of direct combination. In many of the reactions which occur in the atmosphere, such as the formation of tarnish films, the processes are somewhat obscure. When zinc corrodes in the ordinary atmosphere an oxide film forms in the early stages which is pseudomorphic with the metal [11] but which is converted eventually into the ordinary granular form of zinc oxide. The rate of corrosion of zinc is determined by the rate of this conversion

process,[12] the granular oxide film having no retarding influence. This is shown by the linear relationship in Fig. 1 which compares the corrosion-time curves for zinc, copper, lead, and iron in the atmosphere. In this case corrosion is expressed in terms of weight gain due to the accumulation of corrosion products. It will be observed that the relationship for copper is a parabolic one, indicating that the process is controlled by the rate of diffusion of oxygen through the increasingly thicker oxide film. Up to a thickness of about 10 Å the film is invisible and when formed in pure air is impervious to volatile sulfur compounds.

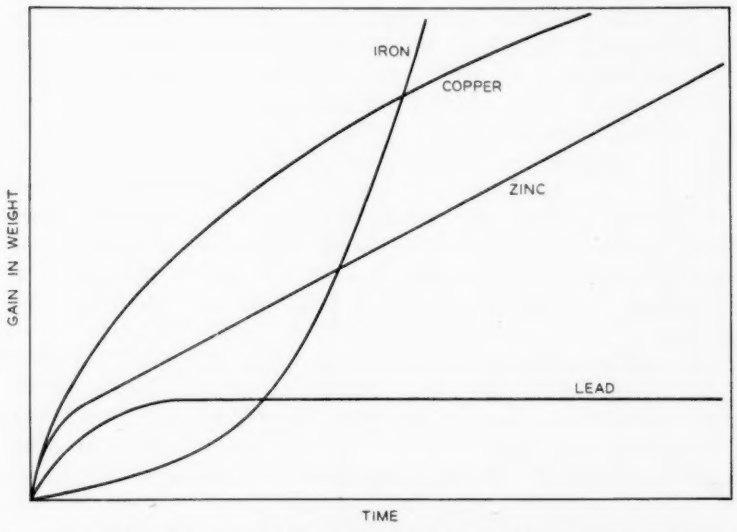


Fig. 1—Corrosion-time relationships characteristic of certain metals exposed to the atmosphere.

A third type of corrosion-time curve represented by lead becomes parallel with the time axis after the initial stages. Evidently the film in this case is impervious to the constituents of the environment. The curve given for iron indicates that the film which forms in the initial stages of the exposure exerts an accelerating influence upon the subsequent rate of oxidation.

It should be mentioned that this acceleration occurs only at humidities above what has been called the "critical" humidity. By this term is meant the relative humidity corresponding to the vapor pressure of a saturated solution of the corrosion products, which depends upon the composition and to some extent the structure of these

products. For iron this is somewhere between 40 and 65 per cent relative humidity, probably nearer the latter figure.[13] The atmospheric corrosion products of the non-ferrous metals are in general less deliquescent, i.e., have higher critical humidities; for nickel it appears to be above 70 per cent relative humidity and for zinc and copper above 75 per cent. It has been suggested that the marked increase in the resistance of copper containing about 0.5 per cent arsenic to atmospheric corrosion is due to the fact that the presence of arsenic renders the corrosion product less hygroscopic.[14] The effect of copper in copper bearing steel may be of the same nature;[15] apparently the inhibiting action in this case does not lie in the production of a film which is any more resistant to attack initially than that on ordinary steel.[16] In dust-free air, even at high humidities, iron does not corrode but forms an invisible protective film. Electron beam studies[17] have indicated the structure of this film to be a form of ferric oxide which has been designated as α Fe_2O_3 in contrast to the composition of a non-protective form which appears to be γ FeO·OH.

The presence of dust particles in the atmosphere greatly increases the rate of corrosion of iron. In this case, as well as in the accelerated attack which occurs above the point of critical humidity, the process involves the displacement of hydrogen from water. Other common examples of this type are the solution of metals in acids or alkalies, the reaction of sodium with water, the deposition of metallic copper from copper sulfate solution by metallic zinc and in general the corrosion of metals in moist air, in soils and in water. It is well established that the process of corrosion in these cases is electrolytic in character, i.e., that corrosion occurs by means of the operation of small galvanic cells at the surface of the metal. The primary reactions of these cells may be and generally are followed by important secondary chemical reactions of the products of electrolysis with the constituents of the environment. Between the anode and cathode areas there is a flow of current through intervening electrolytic paths of greater or lesser resistance. Naturally the amount of corrosion is proportional to the amount of current flow.[18] It is largely the distribution and size of anode areas which determines the character of the corrosive attack. For a given amount of metal dissolution the existence of relatively few anode areas small in size obviously leads to pitting, while if there are numerous anodes uniformly distributed, corrosion will likewise be uniform. The distribution of anodes is determined by the inhomogeneity of the base metal, the character of the films which are formed, the accidental contact of inert bodies and the conductivity of the surrounding electrolytic media.

The potential differences which are responsible for the existence of corrosion cells arise either from some chemical or physical inhomogeneity of the metal or from some inhomogeneity of the environment at the metal surface. These cells or galvanic couples provide the means whereby the natural tendency of metals to corrode may express itself. The nature of the electrodes composing some of these electrolytic cells is given in the following table:

Cathode	*Anode*	*Example of Cell*
Noble metal	Base metal	Two-phase alloys, metal containing metallic impurities, two metals in contact or metal with porous metal coating.
Metal oxide	Metal	Iron with porous oxide scale.
Dust	Metal	Iron with carbonaceous dust particles on surface.
Metal freely exposed to oxygen	Metal in diminished oxygen supply	Exclusion of air at point of contact of metal and inert material or decreased oxygen concentration at bottom of pits on surface of metal.
Metal in concentrated acid, alkali or salt solution	Metal in dilute acid, alkali or salt solution	Metal in contact with solution of two different concentrations or two different solutions.
Annealed or coarsely crystalline metal	Same metal strained or of small crystal size	Metal which has been subject to non-uniform heat treatment or cold working.

It will be observed that there are ample possibilities for a metal of even high purity to corrode. The existence of more than one metallic phase in metals in industrial use often does not have a significant bearing upon their corrosion behavior. For example, studies conducted by these Laboratories have shown that high purity lead, lead hardened with 1 per cent antimony, 3 per cent tin or 0.03 per cent calcium, when used as cable sheathing show approximately the same resistance to corrosion. While one of these materials may be somewhat more corrodible than others in a given natural environment, the reverse will be true for another set of conditions. The environment is of far greater importance than the composition of the sheathing and consequently the control or avoidance of corrosion is attained by maintaining the cable plant in non-corrosive surroundings.

Of the types of corrosion cells listed above, those due to concentration differences in the surrounding medium are among the most prevalent. Differential aeration is one of the most common causes of corrosion. Lead is corroded beneath the point of contact with a large grain of sand when in an atmosphere containing oxygen and water vapor.[19] The point of contact is less accessible to oxygen than surrounding parts. The potential of this cell is somewhat less than 0.1 volt, the value determined in these Laboratories for the difference between the

potentials of lead in identical solutions exposed respectively to oxygen and to an inert gas, nitrogen. An even more striking example of the operation of oxygen concentration cells is observed when glass beads are in contact with a lead surface wet by a film of sodium chloride solution exposed to the air.[20] After a period of a few months a ring of bright red corrosion product forms around an anodic pit located directly beneath each bead. This red corrosion product, a red tetragonal form of litharge, is characteristic of an alkaline attack. Apparently sufficient caustic soda to cause corrosion (a solution of pH 12 approximately) was produced by the differential aeration cell resulting from the contact of the bead with lead.

The foregoing paragraphs have described some of the complexities encountered in corrosion processes. In view of these complexities it has been one object of these Laboratories for some years to advance the development of a generalized theory of corrosion applicable to all cases of corrosion of the replacement type, since it is this type of process which prevails in atmospheric, soil and water exposures. Direct combination of metals with non-metallic elements is limited largely to somewhat extreme conditions such as those of industrial processing. Attention will now be directed to the general theory of corrosion in its present state of development.

The fundamental reactions of corrosion processes of the replacement type as represented by the operation of corrosion cells are as follows:

$$\text{At the anode: } M \xrightarrow{-e} M^+;\ 2(OH)' \xrightarrow{-2e} O + H_2O.$$

That is, the metal sends ions into the solution or there are plated out non-metallic elements such as oxygen. Either process is accompanied by a loss of electrons. If corrosion is induced by an externally applied potential, as for example when stray electrical currents flow from underground metallic structures to earth, oxygen atoms may leave the surface in the form of molecular oxygen.

$$\text{At the cathode: } M^+ \xrightarrow{e} M;\ H^+ \xrightarrow{e} H;\ 2H \rightarrow H_2 \text{ gas.}$$

This process consists in plating out either metal or hydrogen atoms. In the latter case atomic hydrogen either leaves the metal surface as hydrogen molecules or acts as a reducing agent, being in turn oxidized. The electrode reactions may be combined in equations as follows:

1. Solution of the metal: $M + H^+ \rightarrow M^+ + H$.
2. Removal of hydrogen: (*a*) $2H \rightarrow H_2$,
   (*b*) $2H + O \rightarrow H_2O$.

The corrosion reaction represented by the first reaction automatically stops when the metal surface becomes covered with hydrogen atoms and can only proceed when and as hydrogen is removed by one of the processes above given. Some metals evolve molecular hydrogen if the potential of the corrosion cell is slightly over that required to plate out hydrogen atoms. In the greater number of cases, however, molecular hydrogen is not liberated rapidly, i.e., as bubbles, unless the potential is substantially higher than that required to plate out the atomic form. The additional potential required to evolve molecular hydrogen against the normal atmospheric pressure is termed "overvoltage." In the case of metals of high hydrogen overvoltage, oxygen or oxidizing substances are required to facilitate the removal of the hydrogen film if corrosion is to proceed at an appreciable rate. Consequently corrosion may be controlled by the rate at which oxygen reaches the metal surface and "depolarizes" it. In many cases both processes are operative. For example, when iron is corroded in dilute potassium chloride solutions it is interesting to note that under a pressure of one atmosphere of oxygen 1/13 of the total corrosion is accompanied by the discharge of hydrogen gas and 12/13 by the oxidation of hydrogen to form water. When the oxygen pressure on the system is raised to 25 atmospheres by conducting the experiment in a closed bomb the total rate of corrosion is increased 45-fold owing to the increased rate of hydrogen oxidation, the rate of hydrogen evolution being practically unaffected. In practical cases, as would be expected, either the character of the corroding medium or the purity of the metal may affect this ratio of hydrogen control to oxygen control. In tests in which iron specimens were totally immersed in sea water exposed to oxygen, about 35 per cent of the total corrosion was accompanied by hydrogen evolution, whereas for similar specimens in a half-normal solution of pure sodium chloride (which corresponds roughly to the salt content of sea water) only 5.6 per cent of the corrosion was of the hydrogen evolution type.[21] The presence of one part per million of impurity in zinc considerably accelerates the rate of corrosion [22] mainly by stimulation of hydrogen evolution.

The intensity with which a metal tends to send metal ions into solution increases with the basic character of the metal. The presence of ions of the metal in the solution may be considered to constitute a force opposing this ionization tendency and the value of the resulting equilibrium is known as the potential of the metal in that solution. The value is constant or static only so long as there is no flow of current between the metal and solution. The molal potentials and normal potential of metals are their static potential in solutions of

their salts in which the metal ion concentrations are molal and normal respectively. A table of the values of such potentials constitutes the so-called E.M.F. series. This comparison of metals is of little value in predicting either the driving force or the rate of operation of practical corrosion cells, the electrodes and ionic concentrations of which bear little correspondence to those defined in the E.M.F. series. Moreover, the practical case is complicated in many instances by the gas electrode behavior of the metal and by the flow of current during the corrosion process. For example, the potentials of many metals when measured in the atmosphere are much more noble or cathodic than might be expected from a knowledge of the E.M.F. series. This is due to the fact that in the presence of moisture and oxygen, metals may function wholly or in part as oxygen electrodes. The exact values of these electrodes depend upon the concentrations of oxygen present, and upon the acidity of the solution. This explains the origin of the potential difference of the differential aeration cells to which reference has been made previously.

If the corrodibility of copper in the presence of moisture were judged solely from the position of the metal in the E.M.F. series no attack would be expected, since in this series copper is more noble than hydrogen, the element which must be displaced in the corrosion process. As a matter of fact copper does not corrode even in hydrochloric or sulfuric acids in the *absence of available oxygen*. It is readily corrodible, however, in nitric acid because in effect under these circumstances the position of the hydrogen electrode is rendered cathodic to copper (i.e., more noble than copper) owing to the depolarizing influence of oxygen. It is probably for this same reason that oxygen markedly accelerates the corrosion of monel metal in 3 per cent sulfuric acid.[23]

The change in electrode potential with current flow, polarization, may be illustrated in a simple experiment as follows: If the zinc coating is removed from a portion of the surface of a strip of galvanized iron, exposing thereby the underlying iron surface, one has what amounts to a simple galvanic zinc-iron couple. If this couple is completely immersed in a dilute salt solution and potential measurements are made at different points on the iron and zinc surfaces by means of a calomel half-cell, it will be observed that the potentials of iron and zinc at some distance from the iron-zinc interface are approximately those values obtained for these metals separately in the same electrolyte; while, on the other hand, the value of the iron potential near the interface has become more anodic and the potential of zinc near the interface has moved in the cathodic direction, i.e., the difference in potential between iron and zinc near the interface of the two metals is appreci-

ably less than when taken at points more widely separated. Figure 2 gives a schematic representation of this experiment. The change in potential is the result of current flow through the electrolyte from zinc to iron. The current densities are highest in the region of the interface, the metal ion concentration becoming increased at the anode area and decreased at the cathode area, producing thereby anodic and cathodic polarization, respectively.

This polarization behavior of corrosion cells largely determines the rate of corrosion. It is obvious that the effective potentials of corrosion cells may be reduced by polarization to zero, in which case the rate of corrosion is limited to that required to maintain this polariza-

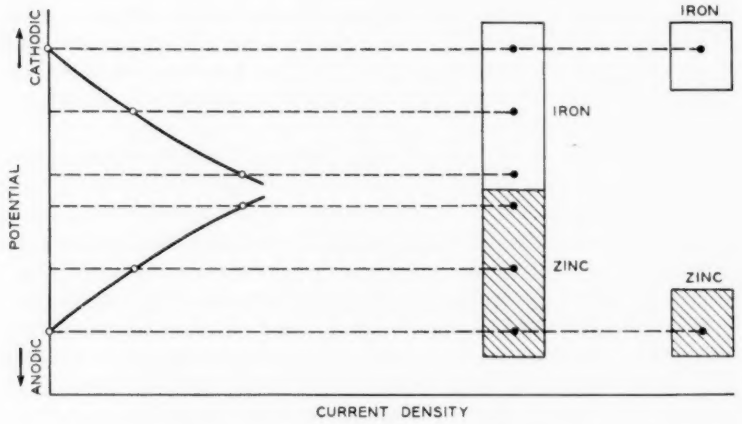


Fig. 2—Illustration of galvanic polarization.

tion. In other words, the progress of corrosion may be controlled by the extent either of anode polarization or cathode polarization or both, that is, either one may determine the final result. Figure 3 represents the variety of current density-potential relationships which may exist in corrosion cells. In Cell 1, in which there is no appreciable polarization of either anodic or cathodic areas (as indicated by the small change of potential with current), corrosion current flow is limited by the resistance of the electrolytic paths between anodes and cathodes and since this may be small if these areas are contiguous the corrosion rate may be high. In Cell 2 the anode is highly polarized as represented by the solid line or progressively less polarized as the point of intersection with the non-polarized cathode occurs at higher and higher current densities as represented by the dotted lines. In a

similar way Cell 3 shows cathode polarization only and Cell 4 both anode and cathode polarization. Since the rate of corrosion is proportional to the flow of current per unit area it is obviously limited in the last three cases by the values of current density at which the polarization curves intersect. In the presence of adequate oxygen or in cases where hydrogen is readily discharged corrosion cells are likely to resemble Cell 1. Where this is not the case, as in the absence of oxygen or where the cathodes have high values for hydrogen overvoltage, the result will be as shown for the cathodically polarized Cell 3. The presence of an inhibitor such as a positively charged colloid or

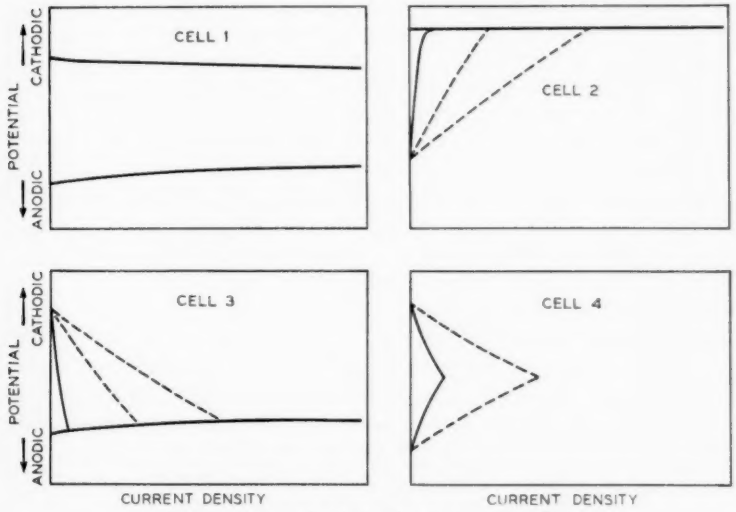


Fig. 3—Types of polarization in corrosion cells.

the amalgamation of the metal surface with mercury are other conditions which promote cathode polarization. On the other hand, the action of passivating agents such as chromates, silicates and in some cases sulfates, carbonates, etc., is to produce anodic polarization as in Cell 2. It will be observed that whereas in the presence of inhibitors of the type mentioned above which polarize the cathode the resulting potential of the corrosion cells and therefore of the metal specimen as a whole should move in the anodic direction as the process of inhibition takes place, in the case of passivating agents (which influence anode processes) the effect of increasing passivation is a trend of potential in the cathodic or noble direction. In both cases corrosion is retarded or prevented entirely.

The manner in which the conductance of the surrounding electrolyte influences the rate of corrosion is illustrated in Fig. 4A in which the upper curve represents the cathodic and the lower the anodic polarization. Assuming equal anodic and cathodic areas the corroding current density for the lower conducting solution is represented by "M" and that for the higher conducting solution by "N." In the actual case where anodes and cathodes are in close juxtaposition, the internal resistance is low and consequently the corroding current density approaches that represented by the intersection of the polarization curves.

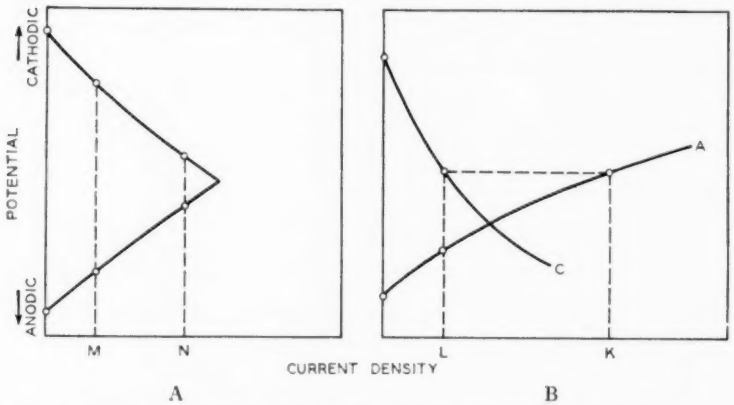


Fig. 4—Effect of conductance and of electrode area on corrosion current densities.
M = Lower conducting solutions.
N = Higher conducting solutions.
L = Corrosion current density for cells of equal cathode and anode areas.
K = Corrosion current density when ratio anode area to cathode is small.

Thus far consideration has been confined for the sake of simplicity to corrosion cells in which the anodic and cathodic areas are equal. Usually in actual experience this is not the case. In corrosion characterized by pitting, the anodic area is generally small compared to the cathodic areas. This situation is illustrated in Fig. 4B in which it will be seen that under these conditions a high corroding current density corresponding to a rapid rate of attack may occur. Conversely, in cases where the ratio of anode areas to cathode area is large, the rate of attack will be small, being thus controlled by cathodic polarization. In this connection it may be of interest to consider the effect of impurities upon rate of corrosion. If the contaminating metal is anodic and exists as a separate phase it will tend to dissolve with the formation of small pits which having once formed may possibly

continue to function as the anodes of oxygen concentration cells. If, on the other hand, the metallic impurity is cathodic and present as a separate phase, corrosion will be rather more uniform in character and its rate will be controlled in the absence of oxygen by the ability of the impurity to discharge hydrogen. Unless its overvoltage is low, that is, unless it discharges hydrogen readily, the rate of corrosion will be slow, the corrosion cells being polarized cathodically. The presence of oxygen or oxidizing agents under these conditions will depolarize these cathodic areas and accelerate corrosion.

From the foregoing it is apparent that a knowledge of the anodic and cathodic current density potential relationships which are established on the surface of a metal in a given environment would make possible an understanding of the processes which are taking place and lead to a prediction of corrosion behavior. It is generally impossible to measure these quantities as they relate to individual corrosion cells owing to a lack of knowledge of the electrode areas involved. Probably these comprise a wide range of sizes and change in size with the progress of corrosion. Sometimes the nature of the cathodes is also uncertain. Practically, however, it is a simple matter to determine a composite of the resultant potentials and their change with time. These time-potential measurements indicate whether the process is anodically or cathodically controlled and in some cases furnish information as to the rate at which it is proceeding, experimental facts which are of value in predicting corrodibility. A recording potentiometer is of considerable assistance in this connection.

Figure 5 illustrates schematically the correlation between these time-potential relationships and the anodic and cathodic polarizations which determine their positions. It will be seen that the potential of iron in a solution of potassium sulfate (represented by the lower solid curve) is mainly determined by the anodic potential of iron in the solution, the cathodic areas being polarized. When potassium chromate is added to the solution the resultant potential of iron is shifted markedly in the cathodic direction, the position being determined by anodic polarization. The actual values of the potential of iron in these cases are of the same order as that of iron alloyed and rendered passive by the addition of chromium and nickel.[24]

In a solution containing hydrogen peroxide, iron is passive even in the acid range as shown in Fig. 6. The abrupt cathodic shift in the potential of iron in the region of pH 6.5–6.8 also shown in Fig. 6 is due evidently to a film which affects both anodic and cathodic areas and which judged from this position would be expected to be less protective than the more pronouncedly passive films (indicated by

their cathodic positions) produced by hydrogen peroxide and chromate ions. A trend of potential in the anodic direction with time, while it may suggest the breakdown of a passive film, does not necessarily indicate the onset of a corrosive attack, for should there also be present substances such as positively charged colloids or other products [25] which tend to raise and maintain hydrogen polarization at the cathodic areas, the metal may suffer little or no attack.

Whether the films which form in corrosion processes are protective in character depends to a considerable extent upon their position or location with reference to the surface of the metal and this in turn

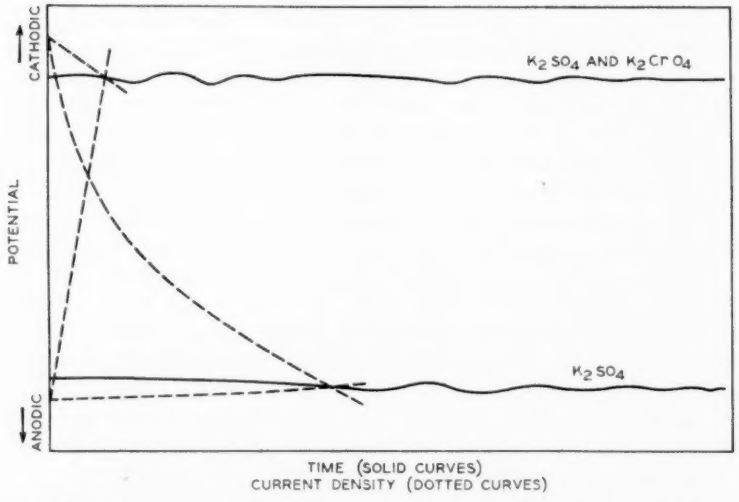


Fig. 5—Time-potential relationship of iron in K_2SO_4 as affected by addition of K_2CrO_4.

depends upon the solubility of the corrosion products in the medium adjacent to the surface. When iron corrodes in the presence of moisture, ferrous ions are produced at anodic areas and hydroxyl ions at cathodic areas, the process continuing until the solubility limit of ferrous hydroxide is attained, whereupon this compound begins to be precipitated as a gelatinous film over the surface of the metal. Increasing the alkalinity of the environment naturally represses the solubility of this compound, precipitating it with less solution of iron. In the absence of oxygen the ferrous hydroxide film tends to inhibit corrosion by maintaining hydrogen polarization. When, on the other hand, oxygen is accessible to the system, ferrous ions are oxidized with the result that ferric hydroxide being less soluble than ferrous

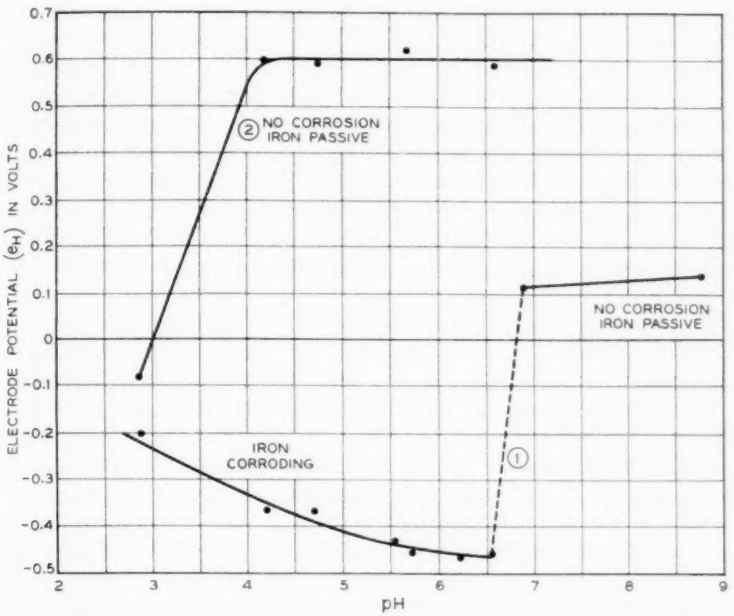


Fig. 6—Effect of acidity and of hydrogen peroxide on the equilibrium potential of iron in buffer solutions of varying acidity.

1. Potential of iron in tenth normal solutions of sodium acetate-acetic acid solutions.
2. Potential of iron in tenth normal solutions of sodium acetate-acetic acid solutions + 0.6 per cent hydrogen peroxide.

hydroxide is precipitated and forms a more or less porous rust film at an appreciable distance from the metal surface. Owing to the mildly amphoteric nature of iron there may exist, especially under alkaline conditions or higher temperatures a considerable concentration of ferrite (FeO_2') ions which, upon reacting with ferrous ions, may precipitate ferrous ferrite (Fe_3O_4) or black magnetic oxide of iron which, also being precipitated in a somewhat granular form at some distance from the surface of the metal, is non-protective. In contrast to these examples is the highly protective film of silicate which presumably forms upon lead and lead rich alloys when immersed in water or soil solutions containing as little as ten parts per million of silicate. As is well known, distilled water is corrosive to these metallic materials. Were it not for this fortunate effect of silicates upon lead it is doubtful that it or its alloys could be used for cable sheathing in the present type of underground construction which permits exposure to soil and surface waters at times.

Studies have shown that the points of failure of air-formed films on iron and steel surfaces indicated by the initial appearance of anodic or rust spots depends upon the previous history of the specimen and upon the medium in which the test is conducted.[26] For example, an increase in time of pre-exposure to oxygen or exposure to higher temperatures decreases the number of initial anodes, while increasing the chloride content of the medium or the presence of sulfide on the metal surface [27] increases them. Whether corrosion continues at the points of initial attack often depends upon the self-healing ability of the film, that is, upon plugging the fissures or pores in the film with corrosion products.

Various methods have been considered for the determination of the quality of protective films. In the case of aluminum and its alloys the amount of leakage current which may pass through anodically formed films throws some light upon resistance to corrosion. Another promising method applied to iron steel and alloy steels has been to determine by potential measurement the amount of chloride required to destroy passive films formed in water or chromate solutions.[28]

The rate of film formation is in a sense a measure of the activity of a metal surface, that is, a measure of the rate at which a metal might corrode in a homogeneous environment in the absence of film forming constituents. For example, it will be seen in Fig. 7 that the potential

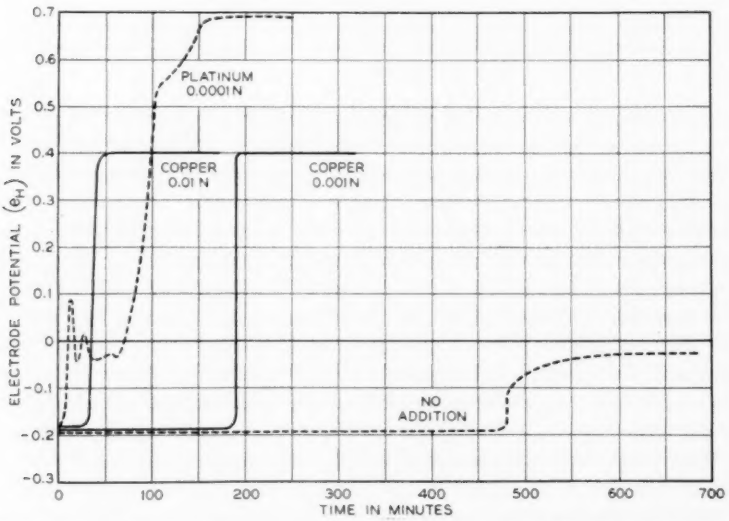


Fig. 7—Effect of traces of copper and platinum on the potential of lead in 0.1N H_2SO_4.

of lead in tenth normal sulfuric acid breaks rather abruptly after about 500 minutes to the potential of a gas electrode. Apparently the anodic areas become progressively covered with a film or sulfate until substantially the entire surface is passive. Upon the introduction of a drop of thousandth normal copper sulfate, the sulfation of a similar lead specimen is consumated in about 200 minutes at the end of which time the potential breaks to that of the cathodic areas of copper which have been formed by replacement deposition. A still higher concentration of copper brings about sulfation still more rapidly and when the solution is contaminated with platinum a break to the potential of platinum occurs after a still shorter period. In the same manner, the relative corrodibilities of leads of various purities and certain lead alloys [29] has been compared.

The foregoing discussion of the application of electrochemical methods to corrosion investigations outlines techniques by means of which it is possible to get information of the following kind. By the position of the potential of a metal against its environment and the trend of this potential with time it is possible to determine whether the corrosion process is controlled by reactions occurring at the anodic areas, the cathodic areas or both, that is, whether there is a tendency toward passivity, inhibition or progressive attack. Measurements of film stability whether in terms of the leakage current which may be passed through the film, or in terms of the amount of film-forming material (such as chromates) required to produce passivity or the amount of film destroying material (chlorides) required to render the metal active, furnish information as to the quality of corrosion resistant films. Finally measurements of the rate at which a film forms on a metal when placed in a film-forming environment also throws light upon the relative surface reactivity of the metal. Such information is of assistance in determining the rates of corrosion in homogeneous corrosive environments or the rate of passivation in film-forming environments. It is evident in all of these cases that the interpretation of the experimental data which are obtained and the application of the findings to practical corrosion problems is considerably facilitated by a chemical knowledge of the environments in which metals are used as well as the composition and physical state or structure of the metallic material. With such measurements and such knowledge it is possible to predict corrosion behavior and to obtain an understanding of corrosion problems usually not possible by ordinary empirical corrosion tests.

To summarize, the process of corrosion may be one of direct combination of a metal and a non-metal or it may be one in which hydrogen

or another metal is displaced from the medium at the surface of the metal. In either case reaction products appear and usually exert a controlling influence upon the progress of attack. In the replacement type of corrosion process, in which the attack occurs by means of the operation of small galvanic couples at the surface of the metal, it is possible to consider separately those influences which affect anode behavior and those involved in cathode behavior. The course of corrosion or resistance to corrosion may be explained in terms of the anode or cathode control of the process. It is apparent then that a knowledge of both the composition and condition of a metal surface and of the surrounding environment is requisite to an understanding of corrosion problems.

References

1. Speller, *Tech. Pub. No.* 553C, *A. I. M. E.*, February 1934.
2. Hudson, *Trans. Faraday Soc., 25,* 177 (1929).
3. Logan and Taylor, *Bur. Stds. Jour. Research, 12,* 119 (1934).
4. Speller, loc. cit.
5. Bengough, *Jour. Soc. Chem. Ind., 52,* 195 (1933).
6. Moore and Liddiard, *Jour. Soc. Chem. Ind., 54,* 786T (1935).
7. Patterson and Hebbs, *Trans. Faraday Soc., 27,* 277 (1931).
8. Logan, *Trans. Electrochem. Soc., 64,* 113 (1933); Ewing, *Am. Gas. Assn.* Report on Pipe Coatings and Corrosion (1933); Dennison, *Bur. Stds. Jour. Research, 7,* 631 (1931); *Am. Gas Jour.*, Sept. 29, 1932; Dennison and Hobbs, *Bur. Stds. Jour. Research, 13,* 125 (1934).
9. Forrest, Roetheli and Brown, *Ind. & Eng. Chem., 23,* 650 (1931).
10. Johnston, *Ind. & Eng. Chem., 26,* 1238 (1934).
11. Finch and Quarrell, *Proc. Phys. Soc., 46,* 148 (1934).
12. Finch, *Discussion Trans. Faraday Soc., 31,* 1116 (1935).
13. Vernon, *Trans. Electrochem. Soc., 64,* 31 (1933).
14. Vernon, *Trans. Faraday Soc., 27,* 225 (1931).
15. Borgmann, *Bell Lab. Record, 10,* 233 (1932).
16. Mears, Sec. D Part 2, *Third Report of Corrosion Comm. of Iron and Steel Inst. and British Iron and Steel Federation* (1935).
17. Rupp, *Kolloid-Z., 69,* 369 (1934).
18. Evans, Bannister and Britton, *Proc. Roy. Soc., A131,* 355 (1931).
19. Burns and Salley, *Ind. and Eng. Chem., 22,* 293 (1930).
20. Experiments carried out by Mr. H. S. Phelps of the Philadelphia Electric Company, through whose courtesy the author was permitted to inspect the results.
21. Bengough, *Jour. Soc. Chem. Ind., 52,* 229T (1933).
22. Harris, *Trans. Am. Electrochem. Soc., 57,* 313 (1930).
23. Searle and Worthington, *Int. Nickel Co. Bull.* T-3.
24. Bannister and Evans. *Jour. Chem. Soc.* 1361, June 1930.
25. Herzog, *Trans. Electrochem. Soc., 64,* 87 (1933).
26. Mears and Evans, *Trans. Faraday Soc., 31,* 527 (1935).
27. Mears, loc. cit.
28. Fenwick, *Indus. and Engg. Chem., 27,* 1095 (1935).
29. Haring and Thomas, *Trans. Electrochem. Soc. Preprint* 68–13 (1935).

Magnetic Measurements at Low Flux Densities Using the Alternating Current Bridge

By VICTOR E. LEGG

A resumé is given of the basic relations between the magnetic characteristics of the core of a coil and the inductance and resistance of the coil as measured on an alternating current bridge. Modifications of the simple relations to take account of the interactions of eddy currents and hysteresis in the core material are developed, and are seen to require a more complicated interpretation of the data in order to obtain an accurate separation of the eddy current, hysteresis, and "residual" losses. Means are described of minimizing or eliminating the disturbing effects of distributed capacitance, leakance and eddy current loss in the coil windings. Essential details of the alternating current bridge and associated apparatus, and of the core structure, are given.

THE modern alternating current bridge, with its high precision and sensitive balance, has almost completely superseded the ballistic galvanometer for determining the magnetic properties of core materials at the low flux densities employed in telephone and radio apparatus. The suitability of the alternating current bridge for this purpose has been recognized for some time,[1] but the continued improvements in magnetic materials, and the more exacting requirements of modern communication apparatus, have necessitated refinements in apparatus, in technique, and in interpretation of measurements. This paper considers the modified technique required to take account of eddy current shielding and hysteresis in the magnetic core, distributed capacitance and leakance in the coil winding, and the necessary details of the bridge and associated apparatus to realize the desired accuracy of measurements.

Fundamentally, the a-c. method involves measurements of the inductance and effective resistance of a winding on the test specimen, such measurements being made at several frequencies, and at several values of current.[2] From these measurements the magnetic properties of the test core can be computed for the low flux density range. The details of such calculations will be given below, beginning with approximate methods, and proceeding to successively more accurate computations.

[1] M. Wien, *Ann. d. Physik* [3] **66**, 859 (1898).

[2] The annular form of magnetic core, wound with a uniformly distributed test winding will here be treated, but the results will be found to be readily transferable to other forms of core.

Simple Analysis of Inductance; Hysteresis Neglected

The magnetizing force in a thin annular core of mean diameter d (cm.) due to current i (ampere) flowing in a uniform winding of N turns is

$$H = \frac{0.4Ni}{d} \text{ œrsted.} \tag{1}$$

In a core of appreciable radial thickness, the effective magnetic diameter rather than the arithmetical mean diameter must be used in this and following equations, as will be explained in eq. (61).

When the bridge is balanced with sinusoidal current of peak value i_m, the peak inductive voltage drop across the standard coil, $2\pi fLi_m$, must equal that across the test coil $2\pi fN\Phi_m \times 10^{-8}$, where Φ_m is the peak magnetic flux in the coil.[3] Whence, for an annular coil,

$$L = \frac{4N^2\Phi_m}{H_m d} \times 10^{-9} \text{ henry.} \tag{2}$$

The flux within an annular coil is composed of that in the core and that in the air space. The expression for inductance can therefore be separated into two terms, giving

$$L = \frac{4N^2}{d}(\mu_m A + A_a) \times 10^{-9} \text{ henry,} \tag{3}$$

where A and A_a are the cross-sectional areas of the core and residual air space, respectively, and μ_m is the magnetic permeability of the core, now assumed to be constant throughout the cycle.

The inductance due to the core alone is then

$$L_m = L - L_a' = \frac{4N^2\mu_m A}{d} \times 10^{-9}, \quad \text{where} \quad L_a' = \frac{4N^2A_a}{d} \times 10^{-9}. \tag{4}$$

The permeability of the core material can be obtained from this inductance as

$$\mu_m = \frac{L_m d}{4N^2 A} \times 10^9. \tag{5}$$

The peak flux density in the core is derived from eqs. (5) and (1) as

$$B_m = \mu_m H_m = \frac{\sqrt{2}L_m I}{NA} \times 10^8 \text{ gauss,} \tag{6}$$

where I is the r.m.s. current in the winding.

[3] A list of most frequently used symbols will be found in the appendix.

Simple Analysis of Eddy Current Resistance

Again, at bridge balance, the resistance of the standard is equated to the resistance of the test coil, which is composed of the copper resistance R_c and a resistance which corresponds to the a-c. power P dissipated in the core. Thus

$$R = R_c + P/I^2. \tag{7}$$

Power is dissipated in the core through eddy currents and magnetic hysteresis. Although both types of magnetic loss occur simultaneously, they will first be considered as if occurring alone, after which the details of separating and identifying the two types will be discussed.

The resistance due to eddy current power loss depends upon the form of the magnetic core—whether of laminations, wire, or powder—upon the frequency, upon the permeability of the magnetic material, and upon the hysteresis loss, since this modifies the permeability. It is determined with sufficient accuracy for many practical purposes by calculating the eddy current power loss in a volume element consisting of a thin tube so drawn that neither magnetic flux nor eddy currents cross its surfaces, when the flux it encloses varies sinusoidally, and then integrating between proper limits to include the entire cross-section of the lamination. By this method [4] it can be shown that the power consumption per unit volume of sheet core material is

$$P_1 = \frac{\pi^2 t^2 f^2 B_m^2}{6\rho} \times 10^{-7} \text{ watt}, \tag{8}$$

where t is the sheet thickness in cm., f is the frequency, and ρ is the resistivity of the material in e.m.u.

This relation is derived on the assumption of a very extensive plane sheet with magnetizing force parallel to its surface, but it applies sufficiently well to any sheet material, flat or curved, provided that the magnetizing force is parallel to its surface, and provided that the width of the magnetic sheet is large in comparison to its thickness. These conditions can be fulfilled in a core built up of ring shaped laminations, and wound with an annular winding.

The total eddy current power loss in a core of volume πAd is then

$$P_e = \frac{\pi^3 t^2 f^2 B_m^2 Ad}{6\rho} \times 10^{-7}. \tag{9}$$

As already mentioned (eq. 7), such a power loss in the core of a coil

[4] C. P. Steinmetz, "A. C. Phenomena," p. 195 (1908).

appears in an a-c. bridge measurement as a resistance

$$R_e = \frac{P_e}{I^2} = \frac{\pi^3 t^2 f^2 B_m^2 A d}{6\rho I^2} \times 10^{-7}. \tag{10}$$

Substituting for B_m from eq. (6), and for L_m from eq. (4), gives

$$R_e = \frac{4\pi^3 t^2}{3\rho} \mu_m L_m f^2 \text{ ohm.} \tag{11}$$

Since this equation contains explicitly no geometrical details of the core, other than the sheet thickness t, it is applicable to any type of core in which the flux density is uniform, as it is in an annular core. If the resistivity is expressed in microhm-cm., the eddy current resistance becomes

$$R_e = \frac{0.0413 t^2}{\rho_1} \mu_m L_m f^2 \text{ ohm.} \tag{12}$$

A similar solution for the case of a core consisting of a hank or bundle of insulated magnetic wires of diameter t cm. gives

$$R_e = \frac{\pi^3 t^2}{2\rho} \mu_m L_m f^2, \tag{13}$$

or with ρ in microhm-cm.,

$$R_e = \frac{0.0155 t}{\rho_1} \mu_m L_m f^2. \tag{14}$$

A compressed magnetic dust core can be idealized as composed of closely packed insulated spheres. Although there is considerable concentration of flux at various points in any practical core, the power loss in a sphere of diameter t_1 can be calculated to a first approximation by assuming it permeated by a uniform flux density parallel to the direction of the magnetizing force. Computing the eddy current power loss in a cylindrical shell of such a sphere with shell axis parallel to H, and integrating to obtain the total loss gives

$$P = \frac{\pi^3 t_1^5 f^2 B_m^2}{120\rho} \times 10^{-7}. \tag{15}$$

The power expended in a cubic centimeter of such a core is then

$$P_1 = \frac{\pi^2 t^2 f^2 B_m^2 r}{20\rho} \times 10^{-7}, \tag{16}$$

where t^2 is the mean square sphere diameter, and r is the packing

factor, i.e., the ratio of the volume occupied by metal to the total volume of the core.

The flux density in the spheres is larger than the apparent flux density by the factor $r^{-2/3}$. The eddy current resistance for such a structure then becomes

$$R_e = \frac{2\pi^3 t^2}{5\rho r^{\frac{1}{3}}} \mu_m L_m f^2, \tag{17}$$

where μ_m is the permeability of the core, as calculated from the inductance L_m.[5] With ρ in microhm-cm., this equation becomes

$$R_e = \frac{0.0124 t^2}{\rho_1 r^{\frac{1}{3}}} \mu_m L_m f^2. \tag{18}$$

In cases where merely comparative tests are to be made, or where ρ and t are not known, it is convenient to lump the coefficient of the eddy current resistance in the form

$$R_e = e\mu_m L_m f^2. \tag{19}$$

Simple Analysis of Hysteresis Resistance

In addition to the power loss due to eddy currents, there is a loss caused by magnetic hysteresis. The energy in ergs dissipated per cubic centimeter of core during one hysteresis cycle is

$$W = \frac{1}{4\pi} \oint HdB = \frac{a_1}{4\pi}, \tag{20}$$

where a_1 is the area of the hysteresis loop in gauss-œrsteds. The power consumption on this account in an annular core of volume πAd, carried through f cycles per second, is

$$P_h = W\pi dAf = \tfrac{1}{4} a_1 dAf \times 10^{-7} \text{ watt}. \tag{21}$$

In the same manner as above, this power is observed by an a-c. bridge balanced for the frequency f as a resistance

$$R_h = \frac{8\pi W}{B_m^2} \mu_m L_m f = \frac{2a_1}{B_m^2} \mu_m L_m f \text{ ohm}. \tag{22}$$

By this relation, the hysteresis resistance can be used to compute the energy loss per cycle W, or the hysteresis loop area a_1, which would be obtained by ballistic galvanometer measurements of sufficient sensitivity.

[5] Cf. R. Gans, *Phys. Zeit.* **24**, 232 (1923).

The Rayleigh Hysteresis Loop

The hysteresis loop area for magnetic cycles at low flux densities (i.e. for which μ_m is not more than 10–20 per cent higher than μ_0) can be calculated from the general shape of such loops. Rayleigh found experimentally [6] that the two branches of such loops are parabolas, that the permeability corresponding to the tips of the loop increases in proportion to the peak magnetizing force, thus,

$$\mu_m = \mu_0 + \alpha H_m, \tag{23}$$

and that the remanent flux is

$$B_r = \frac{\alpha}{2} H_m^2. \tag{24}$$

The loop equation which satisfies the above conditions is

$$B = (\mu_0 + \alpha H_m)H \pm \frac{\alpha}{2}(H_m^2 - H^2), \tag{25}$$

where the points on the upper branch are obtained by using the + sign and those on the lower branch by using the − sign. Recent ballistic galvanometer measurements of high precision on an iron dust core tend to confirm the reliability of the Rayleigh loop equation for low flux densities.[7]

Integrating HdB around the cycle gives an area $4\alpha H_m^3/3$, which can be used in equation (22) to obtain the hysteresis resistance, as

$$R_h = \frac{8\alpha H_m}{3\mu_m} L_m f. \tag{26}$$

Defining the permeability variation with flux as

$$\lambda = \frac{\mu_m - \mu_0}{\mu_0 B_m}, \tag{27}$$

the value of α will be

$$\alpha = \mu_0 \mu_m \lambda, \tag{28}$$

and the hysteresis resistance becomes

$$R_h = \frac{8}{3} \lambda H_m \mu_0 L_m f. \tag{29}$$

[6] *Phil. Mag.* [5] **23**, 225 (1887).
[7] W. B. Ellwood, *Physics* **6**, 215 (1935).

It thus appears that the Rayleigh hysteresis loop implies a definite relationship between the variation of permeability with H or B, as calculated from bridge measurements of inductance at various coil currents, and the observed hysteresis resistance. Efforts to judge as to the general applicability of the Rayleigh form of loop by means of such a-c. bridge comparisons have indicated fairly good agreement for most materials, but occasional deviations as high as 40 per cent.[8] Best agreement is generally found in well annealed and unstressed materials, while deviations are found in such materials as compressed dust cores. In such comparisons, the anomalous residual loss, variously termed magnetic viscosity, and after-effect, is excluded. This additional loss will be discussed below.

Mutual Effect of Rayleigh Hysteresis and Eddy Current Shielding in Sheet Material

Taking the above equation as the simplest general representation of hysteresis loops at low flux densities, it now becomes necessary to review the previous work with additional refinements to include the effects of hysteresis upon eddy currents, and of eddy currents upon themselves, and upon hysteresis. Thus the fact that B varies according to a hysteresis loop equation rather than directly with H modifies the eddy current loss somewhat. Also, eddy currents set up magnetizing forces within the magnetic material which more or less neutralize that applied by the coil winding, and thus effectively shield the inner parts of magnetic laminations of wires. Such eddy current shielding reduces the total flux in the core, thus decreasing the inductance and loss resistance observed at higher frequencies.

The fundamental differential equation giving the relation between B and H at a point x distant from the median plane of a magnetic sheet is[9]

$$\frac{\partial B}{\partial t} = \frac{\rho}{4\pi} \frac{\partial^2 H}{\partial x^2} . \tag{30}$$

For the simple case of constant permeability in which $B = \mu_0 H$, this equation has been solved by Heaviside, J. J. Thomson,[10] and others.

For the case in which B is given by Rayleigh's equation (25), the solution is very much involved. The variable permeability gives rise

[8] E. Peterson, *B. S. T. J.* **7**, 775 (1928).
[9] *E.g.*, Russell, "Alternating Currents," Vol. I, p. 487 (1914).
[10] *Electrician* **28**, 599 (1892).

to odd harmonic voltages which are important from the standpoint of modulation and noise. In a-c. bridge measurements where the balance is made so as to bring the voltages of fundamental frequency to equality, it suffices to carry through the mathematics for this frequency alone. This has been done for sheet and wire cores to an accuracy sufficient for most purposes by W. Cauer.[11] From his results for the inductance and power loss in a laminated core, the apparent permeability and loss resistance are calculated to be

$$\mu_{fm} = \frac{\mu_0}{\theta}\frac{\sinh\theta + \sin\theta}{\cosh\theta + \cos\theta} + \alpha H_m\left(1 - \frac{4\theta^2}{9\pi} - \frac{7\theta^4}{60} + \frac{2\theta^6}{45\pi} + \cdots\right) \quad (31)$$

$$= \mu_0\left(1 - \frac{\theta^4}{30} + \frac{\theta^8}{732} - \cdots\right) + \alpha H_m\left(1 - \frac{4\theta^2}{9\pi} - \frac{7\theta^4}{60} + \frac{2\theta^6}{45\pi}\cdots\right) \quad (32)$$

$$= \mu_0\left[1 + \lambda B_m - \frac{4\lambda B_m\theta^2}{9\pi} - \left(1 + \frac{7}{2}\lambda B_m\right)\frac{\theta^4}{30} + \frac{2\lambda B_m\theta^6}{45\pi}\cdots\right], \quad (33)$$

$$R_{fm} = \frac{2\pi f L_0}{\theta}\frac{\sinh\theta - \sin\theta}{\cosh\theta + \cos\theta} + \frac{8\alpha H_m f L_0}{3\mu_0}\left(1 + \frac{\pi\theta^2}{4} - \frac{7\theta^4}{60} - \frac{\pi\theta^6}{40} + \cdots\right) \quad (34)$$

$$= \frac{\pi f L_0\theta^2}{3}\left(1 - \frac{17\theta^4}{420} + \frac{\theta^8}{600} - \cdots\right) + \frac{8}{3}\lambda B_m L_0 f\left(1 + \frac{\pi\theta^2}{4} - \frac{7\theta^4}{60} - \frac{\pi\theta^6}{40}\cdots\right). \quad (35)$$

The quantity $\theta = 2\pi t\sqrt{\mu_0 f/\rho}$, where ρ is in e.m.u.; and $B_m = \mu_m H_m$, where μ_m is independent of f.

The hyperbolic function parts of these equations are valid at any frequency, but they give only those parts of μ and R which are due to the constant initial permeability μ_0. The series having α or λ as coefficients give the increases due to hysteresis.

The apparent permeability μ_{fm}, which is calculated from the measured inductance, decreases as the measuring frequency is increased. Furthermore, at higher frequencies, this permeability rises less rapidly with rise in measuring current than it does at low frequencies, and it will actually decline with increasing H at frequencies higher than that necessary to make $\theta > 1.6$, approximately. Thus, for the accurate determination of μ_0, μ_m, λ and α, it is necessary to make measurements at frequencies low enough to suppress these correction terms.

[11] W. Cauer, *Arch. f. Elektrotechnik* **15**, 308 (1925).

The equations for resistance are similarly complicated. The first series gives that part of the eddy current resistance which is due to the constant μ_0. The coefficient of the second series indicates that this series involves the hysteresis resistance. However, terms in the second series which contain the factor f^2 will be recognized as eddy current components introduced by the fact that the permeability has been increased from the value μ_0 by the factor λB_m.

The complicated form and slow convergence of the above equation (35) for resistance make it difficult for use in interpreting a-c. bridge measurements. Considerable simplification is effected by dividing the observed resistance (eq. 35) by the observed inductance (from eq. 33) for each measuring current and frequency. Performing this operation, and rejecting series terms in λB_m higher than the first power, gives

$$\frac{R_{fm}}{L_{fm}} = \frac{4\pi^3 f^2}{3\rho}\mu_m f^2\left[1 - \frac{\theta^4}{140}(1 + 5\lambda B_m) + \cdots\right] + \frac{8}{3}\lambda H_m \mu_0 f\left[1 - \frac{\theta^4}{36}(1 - 5\lambda B_m) + \cdots\right]. \quad (36)$$

The coefficient of the first series is identical with the eddy current expression previously derived (eq. 11), which neglected eddy current shielding and hysteresis. The series itself, which includes these other effects, converges rapidly for $\theta < 1$, provided that the value of λB_m is not carried too high.

The coefficient of the second series is identical with the hysteresis expression derived from Rayleigh's equation (29) in which eddy currents were neglected. The second term of this series gives the amount by which eddy current shielding reduces hysteresis resistance at higher frequencies. It appears to converge less rapidly than the series for the eddy current resistance, but this is partly offset by the decrease of its second term with increase of λB_m. Thus, the coefficients of θ^4 become equal to 1/88 in both series if $\lambda B_m = 13/110$. This value of λB_m is reached when the flux density in the material is large enough to raise the permeability some 10 per cent above μ_0. Evidently, this value of λB_m can be exceeded somewhat without making the coefficients excessively large. However, if the measurements are made at too high flux densities, the hysteresis loops diverge more and more from the simple Rayleigh loop, and the present analysis becomes inapplicable.

In a-c. bridge measurements it is seldom desirable to measure at flux densities which will carry the permeability more than 10 per cent above its initial value. If measurements at higher flux densities are

desired, sufficient sensitivity can generally be obtained by wattmeter or ballistic galvanometer methods.

Graphical Separation of Losses

Since it is generally important to distinguish between types of magnetic losses, methods of analyzing the measurements have been devised to accord with the degree of refinement desired. A fairly simple graphical loss separation method is suitable if magnetic shielding can be ignored. However, it will be seen to lead to the inclusion of an additional term to account for the residual loss. If the effect of eddy current shielding is also to be considered, a more complicated analytical method of separation will be found necessary.

With magnetic shielding negligible, eq. (36) reduces to the form

$$\frac{R_m}{fL_m} = \frac{8}{3}\lambda H_m \mu_0 + \frac{4\pi^3 t^2}{3\rho}\mu_m f \tag{37}$$

$$= \frac{8}{3}\lambda B_m \frac{\mu_0}{\mu_m} + \frac{4\pi^3 t^2}{3\rho}\mu_m f \tag{38}$$

or

$$\frac{R_m}{\mu_m f L_m} = aB_m + ef. \tag{39}$$

The last form of the expression is most suitable for routine testing and design purposes. The hysteresis area constant a will be seen to be intimately related to the hysteresis loop area a_1 previously discussed; thus

$$a = \frac{2a_1}{B_m^3}. \tag{40}$$

Within the limits of applicability of Rayleigh's equation, the following relations also apply:

$$a = \frac{8\lambda\mu_0}{3\mu_m^2} = \frac{8\alpha}{3\mu_m^3}. \tag{41}$$

The losses observed on any test core can be separated graphically [12] by calculating the values of $R_m/\mu_m f L_m$ for a fixed value of H_m at all measuring frequencies, and plotting such values against frequency. The slope of the resulting straight line should then give e, and the intercept aB_m. When this process is repeated for other values of H_m, a series of intercepts will be obtained, all of which would be expected to yield a constant value for the loop area constant a. However, this is frequently found not to be true, but if the several intercepts so

[12] B. Speed and G. W. Elmen, *Trans. A. I. E. E.* **40**, 596 (1921).

obtained be plotted against B_m, they generally fall upon a fairly straight line whose slope is a, and whose intercept on the line $B_m = 0$ is c. The value of a so obtained agrees fairly well in most cases with the value calculated from the permeability variation coefficient λ, which is the justification cited above for the Rayleigh equation. The presence of residual loss necessitates rewriting the loss equation with an additional term—

$$\frac{R_m}{\mu_m f L_m} = aB_m + c + ef. \tag{42}$$

The value of the intercept c however, has no counterpart in the Rayleigh equation. It indicates the presence of a power loss proportional to the frequency, and thus similar to hysteresis, but contrarily proportional to the square of the magnetizing force, instead of to the cube. It is found not to contribute to harmonics or modulation generated by a core material, and might thus be represented by an elliptical increment to the Rayleigh loop.[13]

Residual loss has been ascribed to viscosity or "after-effect" in the core material.[14] The chief obstacle to this explanation is the observed constancy of c over a wide range of frequencies, in contrast to the variation to be expected from ordinary viscosity losses. Residual loss has been ascribed to inhomogeneities in the magnetic material[15] which lead to higher a-c. power losses than expected from the area of the hysteresis loop. This explanation seems promising, but the work to date has been chiefly qualitative, and it has not been shown to yield the required additional loss proportional to H^2. The parallel between this loss and eddy current loss, which is also proportional to H^2, is alluring, but the dependence of eddy current loss upon f^2 has remained a stumbling block. The mechanical dissipation of power through magnetostrictional motions seems also a possible explanation.[16]

Somewhat analogous to the residual loss is the excess eddy current loss generally observed. When the observed value of e is used to calculate the resistivity of a magnetic material, it generally gives a value somewhat smaller than the true resistivity, which indicates that the observed eddy current losses are correspondingly too large. The apparent resistivity so obtained approaches the true resistivity quite closely for well insulated laminations of pure, well annealed materials. It is interesting to note that the residual loss for such well

[13] H. Jordan, *Ann. d. Physik* [5] **21**, 405 (1934).
[14] H. Jordan, *E. N. T.* **1**, 7 (1924); F. Preisach, *Zeit. f. Phys.* **94**, 277 (1935).
[15] L. W. McKeehan and R. M. Bozorth, *Phys. Rev.* [2] **46**, 527 (1934).
[16] For a more thorough discussion see W. B. Ellwood, *Physics* **6**, 215 (1935).

annealed materials is also practically absent. For many materials, however, the apparent resistivity falls to 50–75 per cent of the true resistivity. The increase in eddy current loss thus observed is technically very undesirable since it necessitates rolling laminations considerably thinner than otherwise required, in order to suppress eddy current losses sufficiently.

The cause of extra eddy current losses in laminated material is definitely chargeable to the hard, low permeability surface of the material. The eddy current losses are determined largely by the high interior permeability, and the laminar thickness. The material near the surface conducts large eddy currents induced by interior material of high permeability, but it contributes very little to the average permeability for the entire sheet. Removal of low permeability surface material by etching [17] lowers the eddy current losses and increases the average permeability of the core, so that the apparent resistivity approaches more closely the true d-c. value. Of course, selection of material and proper mechanical working and heat treating technique are most desirable in avoiding at the outset such inhomogeneities, with their resulting excessive losses.

Analytical Separation of Losses

For special investigations where the accuracy of the graphical method of loss separation is not sufficient, it is necessary to return to the unabridged form of eq. (36), and employ an analytical method. For example with sheet material,

$$e = \frac{4\pi^2 t^2}{3\rho} \quad \text{and} \quad \theta^2 = \frac{3}{\pi} e\mu_0 f.$$

Rewriting eq. (36) with these substitutions and with an additional term to provide for the residual loss,

$$\frac{R_{fm}}{fL_{fm}} = aB_m\mu_m\left[1 - \frac{9e^2\mu_0^2 f^2}{36\pi^2}(1 - 5\lambda B_m) + \cdots\right] + c\mu_m + e\mu_m f\left[1 - \frac{9e^2\mu_0^2 f^2}{140\pi^2}(1 + 5\lambda B_m) + \cdots\right]. \quad (43)$$

It should be recalled that R_{fm} and L_{fm} are the core resistance and inductance measured at a definite current and frequency, while μ_m is the permeability of the core measured at the same current, but at a frequency low enough to make eddy current shielding negligible.

[17] Legg, Peterson and Wrathall, U. S. Patent 1,998,840 (1934).

Subtracting the value of R_{fm}/fL_{fm} for frequency f_1 from the corresponding value for frequency f_2, and dividing by the frequency interval $\Delta f = f_2 - f_1$, gives

$$\frac{\Delta\left(\frac{R_{fm}}{fL_{fm}}\right)}{\Delta f} = e\mu_m \left[1 - \frac{9e^2\mu_0^2}{140\pi^2}(1 + 5\lambda B_m)(f_1^2 + f_2^2 + f_1 f_2) - \frac{2}{3\pi^2} e\mu_0 \lambda B_m (1 - 7\lambda B_m)(f_1 + f_2)\cdots\right]. \quad (44)$$

An approximate value for e is sufficient in obtaining the correction terms in this equation. With the precise value of $e\mu_m$ thus obtained, the eddy current term in eq. (43) can be calculated for any frequency and permeability. Subtracting the proper eddy current term for each value of R_{fm}/fL_{fm} gives the hysteresis terms as remainders, which can be further analyzed in their relation to magnetizing force, as in the previous graphical loss separation. Loss separations, made thus precisely, reveal frequency variations of apparent resistivity and of the residual loss constant.

Capacitance, Leakance, and Eddy Current Loss of the Winding

In the discussion thus far, it has been assumed that the measured inductance and resistance of a test coil depend solely upon the core permeability and losses. This assumption must be modified under some conditions, for it is found that the distributed capacitance and leakance of the coil winding act as shunt impedances, which may diminish sufficiently at high frequencies to mask the actual inductance and resistance of the coil. Furthermore, the resistance of the test coil includes an amount corresponding to the power expended by eddy currents in the copper winding itself. It will be shown that such disturbing factors can generally be eliminated, either by modifications in the method of core loss separation, for materials in which eddy current shielding is negligible; or by winding the test core to give an inductance low enough to suppress such disturbing factors, for materials in which eddy current shielding is not negligible.

If the distributed capacitance and leakance can be considered as single lumps, C, and G, in parallel with the coil of inductance L and resistance R, the observed inductance at a frequency corresponding to $\omega = 2\pi f$ is found to be

$$L_{\text{obs.}} = \frac{L(1 - \omega^2 LC) - CR^2}{(1 - \omega^2 LC)^2 + 2GR + G^2(R^2 + \omega^2 L^2) + \omega^2 C^2 R^2}.$$

This simplifies at frequencies well below resonance to

$$L_{obs.} = L(1 + \omega^2 LC). \tag{45}$$

Thus the observed inductance tends to increase at higher frequencies on account of distributed capacitance, in contrast to its tendency to decrease on account of magnetic shielding in the core according to eq. (33). If the inductance L is known from low frequency measurements, and if computations from eq. (33) show that it does not decline appreciably because of eddy current shielding at the measuring frequency, the capacitance can be calculated from the relation

$$\omega^2 LC = \frac{L_{obs.} - L}{L}. \tag{46}$$

Similar complications arise in measuring the resistance of a coil at high frequencies. Under the same assumptions as above, the observed resistance is

$$R_{obs.} = \frac{R + G(R^2 + \omega^2 L^2)}{(1 - \omega^2 LC)^2 + 2GR + G^2(R^2 + \omega^2 L^2) + \omega^2 C^2 R^2}.$$

At moderate frequencies, this reduces to

$$R_{obs.} = (R + G\omega^2 L^2)(1 + 2\omega^2 LC), \tag{47}$$

from which it appears that leakance enters as an important part, and that the capacitance gives twice as large an increment for the resistance as for the inductance. The effect of distributed capacitance can be eliminated by dividing the observed value of resistance by $(1 + 2\omega^2 LC)$, where the correction factor is obtained from eq. (46). Thus

$$\frac{R_{obs.}}{1 + 2\dfrac{L_{obs.} - L}{L}} = R + G\omega^2 L^2. \tag{48}$$

The leakance term $G\omega^2 L^2$ can be eliminated as will be shown below.

The resistance term R includes the desired magnetic core resistance, but it also contains the resistance of the copper coil, which may have a considerable eddy current loss of its own. The copper eddy current loss occurs principally in the lower layers of the winding, which are cut by the alternating magnetic flux set up by the current in the winding. It is similar to the eddy current loss in the core material itself, varying with the square of the frequency, to a first approximation.[18] This loss must, therefore, be eliminated before accurate

[18] Cf. M. Wien, *Ann. d. Phys.* [4] **14**, 1 (1904); S. Butterworth, *Exp. Wireless* **6**, 13 (1929).

determination of the core loss is possible. The eddy current resistance of the copper winding is of the form

$$R_{ce} = e_c L_a f^2, \tag{49}$$

where L_a is the total air inductance of the winding, and e_c is the copper eddy current coefficient.

This eddy current coefficient is inversely proportional to the number of strands in the wire, so that it can be minimized by using wire consisting of many insulated strands. It increases somewhat with the number of layers in the winding. The coefficient may be determined for any type of winding by resistance measurements on an air core coil of dimensions and winding details similar to those of the magnetic core to be tested. Subtracting the eddy current resistance R_{ce} so computed, and the d-c. copper resistance R_c, from eq. (48) gives as the residual resistance

$$\Delta R = \frac{R_{\text{obs.}}}{1 + 2\dfrac{L_{\text{obs.}} - L}{L}} - R_c - e_c L_a f^2$$

$$= \mu_m L_m[(aB_m + c)f + ef^2] + G\omega^2 L^2. \tag{50}$$

This residual resistance consists of the core loss resistance, and an increment due to leakance. The latter can be largely suppressed by the use of low leakance insulating materials, by insuring that the winding is free from moisture, and by making the distributed capacitance as small as possible. Furthermore, it is known from experiments on the electrical conductance of insulating materials at elevated frequencies that the "quality" $Q = \omega C/G$ is practically a constant (C is the capacitance associated with G, — in this case the distributed capacitance). Inserting this value of G in eq. (50) gives

$$\Delta R = \mu_m L_m[(aB_m + c)f + ef^2] + 8\pi^3 CL^2 f^3/Q. \tag{51}$$

Theoretically, the coefficients in this equation can be obtained from resistance measurements taken at three different frequencies. Unavoidable errors in the measurements render such an analysis unreliable, so that it is generally preferable to obtain a larger number of observations, and to determine the coefficients graphically. Dividing by $\mu_m L_m f$, and neglecting the air inductance of the coil, the equation becomes

$$\frac{\Delta R}{\mu_m L_m f} = (aB_m + c) + ef + \frac{8\pi^3 CLf^2}{\mu_m Q}. \tag{52}$$

If the data at lower frequencies are sufficiently reliable, this parabola can be extrapolated to give the zero intercept $(aB_m + c)$ which contains the sought for hysteresis constants of the core. Subtracting the intercept so found, and dividing again by f gives

$$\frac{1}{f}\left\{\frac{\Delta R}{\mu_m L_m f} - (aB_m + c)\right\} = e + \frac{8\pi^3 CLf}{\mu_m Q}. \tag{53}$$

This is the equation of a straight line, when plotted against f. The intercept e is the desired eddy current coefficient for the core material. The slope of this line, S, yields the dielectric quality

$$Q = \frac{8\pi^3 CL}{\mu_m S}. \tag{54}$$

Here C is the distributed capacitance, which can be obtained from eq. (46). This relation is useful in comparing the qualities of various insulating and spacing materials, and in calculating the total losses to be expected in any proposed coil.

Accurate Separation by Limiting Inductance

It appears from the above discussion that magnetic loss separations can be made in spite of interference by distributed capacitance, leakance and eddy current resistance of the coil windings, provided that the interference is not too large, and provided that eddy current shielding in the test core is negligible. When the latter condition is not fulfilled, it becomes necessary to suppress the interference due to capacitance, etc., to negligibly small quantities. This is facilitated by proper technique in applying the windings, but any degree of suppression can be secured by sufficient limitation of the coil inductance, as will appear by reference to eq. (47). Although reduction of the coil inductance by using a winding with few turns is desirable in thus suppressing errors, it is undesirable in that it reduces the core loss resistance (cf. eq. 50) to a value which may be difficult to measure accurately on any available bridge. It is thus necessary to wind the test core to an inductance which will yield the largest possible loss resistance, without exceeding the allowable error from capacitance, leakance, and copper eddy current loss. The value of this maximum allowable inductance is obtained by calculating the inductance required to make the errors due to capacitance, leakance, and copper eddy currents at the highest measuring frequency equal to some tolerable small fraction of the core loss resistance.

A good separation of losses requires measurements at four or more frequencies, up to a point where the eddy current resistance is several times the hysteresis resistance, and at four or more values of measuring current in the useful range. The maximum frequency necessary to make the eddy current resistance mount to p times the hysteresis resistance is

$$f_m = \frac{ph}{e}, \qquad \text{where} \qquad h = (aB_m + c), \tag{55}$$

and the total core loss resistance at this frequency is

$$R_m = \frac{\mu_m L_m h^2 p(p+1)}{e}. \tag{56}$$

At this maximum frequency, the observed resistance is

$$R_{obs.} = (R_c + e_c L_a f^2 + R_m + 8\pi^3 C L^2 f_m^3 Q)(1 + 8\pi^2 L C f_m^2). \tag{57}$$

Since it is desired that the observed resistance indicate directly the d-c. copper and core loss resistance, all other terms in eq. (57) may be considered as errors, to be suppressed to a small fraction q of the core resistance. Setting the total error equal to qR_m, and rejecting errors of higher orders gives

$$qR_m = e_c L_a f_m^2 + 8\pi^2 L C f_m^2 (R_c + R_m + \pi L f_m Q). \tag{58}$$

Substituting the above values for f_m and R_m at the maximum measuring frequency gives a quadratic equation in L (neglecting air inductance), which solves quite accurately to require,

$$L = \frac{e[eq\mu_m(p+1) - 8\pi^2 C R_c p]}{8\pi^2 C p^2 h[\pi Q + \mu_m h(p+1)]} - \frac{e_c L_a p}{eq\mu_m(p+1) - 8\pi^2 C R_c p}. \tag{59}$$

Since it is desirable to use a large inductance for ease in resistance determination, it appears that the copper resistance R_c, the copper eddy current coefficient e_c, and the distributed capacitance C should be made as small as possible, while Q should be large.

As an illustrative example, assume a core material of permeability $\mu_m = 100$, to be measured up to a frequency such that $p = 5$, with an error of not more than 1 per cent at the maximum frequency, i.e., $q = 0.01$. Assume also, $e = 25 \times 10^{-9}$, $h = 0.5 \times 10^{-4}$, $C = 25 \times 10^{-12}$, $Q = 20$, $e_c L_a = 10^{-11}$, $R_c = 2$.

The maximum frequency for measurement will then be $f_m = 10{,}000\sim$. The core should be wound to give an inductance of 5.30 mh. At the

maximum measuring frequency the eddy current resistance will be 1.325 ohms and the hysteresis resistance 0.265 ohm. At the lower end of the frequency range, say at 1000∼ these figures become R_e = 0.0132 ohm and R_h = 0.0265 ohm. This indicates that a bridge will be required for such measurements capable of measuring increments of resistance to an accuracy of about 0.0002 ohm at 1000∼, and about 0.002 at 10,000∼.

Computations of this sort show the high quality of a-c. bridge generally demanded for core loss measurements. Such measurements require equipment and a bridge with a maximum of sensitivity of both a-c. and d-c. balances, and a minimum of losses in standards, pick-up, unbalanced impedance to ground, and variable contact.

Bridges, and Test Procedure

The essential features of bridges suitable for core loss measurements will now be mentioned in general terms. It will be appreciated from the above discussion that the specific range of frequencies, and the required resistance sensitivity must be adapted to the loss characteristics of the magnetic core to be measured.

Although certain types of resonance bridges [19] have advantages for measurements at high frequencies, the most suitable bridge for the usual measurement of magnetic core coils is an equal arm inductance comparison bridge,[20] on which inductances can be measured directly, and on which a-c. and d-c. resistance measurements can be made in prompt succession, to eliminate the effect of gradual temperature changes on the resistances of the bridge and test coil. A suitable circuit is shown in Fig. 1.

Inductance coils for bridge standards should be as stable as possible against frequency and current. Although low effective resistance per unit of inductance is desirable, it is more important for core loss measurements to design such standards for a minimum increase of resistance with frequency, so as to keep calibration corrections small in comparison to the resistance increments to be measured. A satisfactory type of standard coil consists of an air core toroidal form, with a bank winding of finely stranded wire. The bank winding minimizes capacitance effects on the observed inductance and resistance of the coil, and the fine stranding minimizes eddy current losses in the copper; cf. eq. (49). The small residual corrections must finally be included as calibrations when making measurements with the aid of standard coils.

[19] W. J. Shackelton and J. G. Ferguson, *B. S. T. J.* **7**, 82 (1928).
[20] W. J. Shackelton, *B. S. T. J.* **6**, 142 (1927).

The effect of contact resistance is minimized by changing few, or preferably no, contacts between a-c. and d-c. readings. This is facilitated by supplying current to one corner of the bridge through the sliding contact of the slide wire resistance used for fine balancing. This excludes contact resistance errors from resistance determinations in which both a-c. and d-c. balances fall within the range of the slide wire, and thus increases the bridge accuracy for small values of effective resistance. Usual precautions as to clean and positive contacts are sufficient for larger resistance measurements.

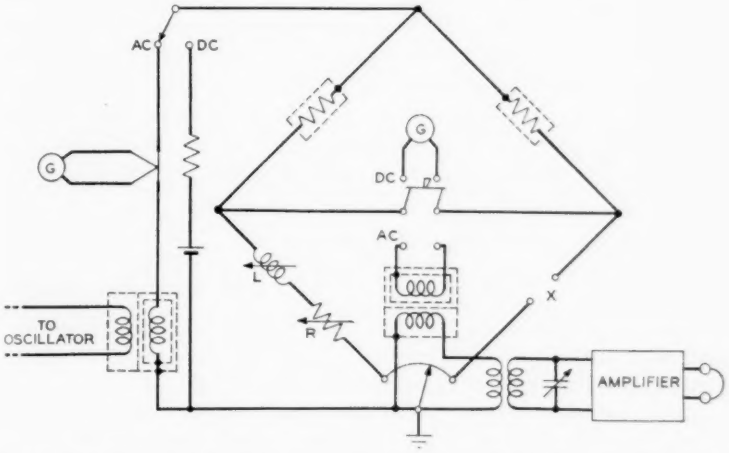


Fig. 1—Diagram of inductance comparison bridge suitable for measurement of magnetic core coils.

The a-c. supply to the bridge should, of course, be a sine wave, and the bridge transformers should be designed for minimum distortion. The frequency should be known accurately, and the voltage should be constant during any set of measurements. Rheostats are required to permit accurate adjustment of bridge current, and they must be designed and shielded to avoid stray coupling with the bridge. A suitable thermocouple is provided for measuring the current into the bridge, from which measurement the current through the test coil can be readily determined, since the bridge has equal ratio arms.

A distortion-free amplifier and a filter circuit tuned to the measuring frequency, are essential for magnifying the bridge unbalance current so as to permit precise measurements. Such unbalances may be detected by a vibration galvanometer at frequencies below say 200~

and by head phones at frequencies within the audible range. For measurements at higher frequencies, a heterodyne detector is needed. Detection with a galvanometer is feasible when used in connection with a rectifier and filter, the filter being required so as to eliminate errors due to currents of extraneous frequencies.

The galvanometer used for d-c. bridge balancing should be sensitive enough to secure resistance readings of precision equal to that of the a-c. balance. The d-c. supply to the bridge should be limited to a current of the same order of magnitude as the a-c. supply, to guard against permanent magnetization of the magnetic material under test.

The usual test procedure is to set the oscillator at the lowest desired frequency, and measure the inductance and resistance of the test coil at several currents beginning at the lowest, increasing to the highest, and returning again to the lowest, so as to detect any tendency for permanent magnetization or magnetic aging. Direct current balances are taken as often as required to keep up with gradual changes of circuit resistance due to room temperature changes, the direction of current through the bridge being reversed each time to detect and eliminate stray currents and thermal e.m.f.'s. The differences between the observed a-c. and d-c. resistances gives the a-c. increment resistance of the test coil, except for corrections on account of the calibration of the bridge and coils. This process is repeated at successively higher frequencies until a suitable range of data has been covered. The resulting data can then be analyzed to show the characteristic of the magnetic core by the appropriate method as described above.

The Test Core

The design of the test core depends upon the physical and magnetic characteristics of the material to be tested. In general the radial thickness should be small in comparison with the diameter. Strain sensitive materials must be protected from mechanical stresses of handling and winding. Types of insulation and winding depend upon the loss characteristics of the core.

In any practical core, the diameter ranges between an inside value d_i and an outside value d_0. Since the diameter enters into the denominator of the expression for H and certain other magnetic quantities, the effective diameter must be calculated and used in such expression rather than the simple mean diameter. The effective magnetic diameter of a core having a rectangular cross-section is

the reciprocal of the value obtained by averaging $1/d$, namely

$$d = \frac{d_0 - d_i}{\log_e \frac{d_0}{d_i}}. \tag{60}$$

This expression can be converted to the following convenient series

$$d = d_m \left[1 - \frac{1}{12} \left(\frac{\Delta d}{d_m} \right)^2 - \frac{1}{180} \left(\frac{\Delta d}{d_m} \right)^4 - \cdots \right], \tag{61}$$

which indicates that the effective diameter is smaller than the arithmetical mean diameter d_m, by an amount depending upon Δd, the difference between inside and outside diameters. This series converges so rapidly that terms beyond the second may be neglected for all practical purposes.

A test core of large radial thickness is to be avoided, when accurate measurements are desired, because of the considerable change of flux density from the inside to the outside diameter, with its accompanying modification of the core permeability. Such variations complicate eddy current and hysteresis behavior, particularly through reaction on the magnetic permeability. If the permeability at every point in the core bears a straight line relationship to the flux density, eq. (27) gives

$$\mu = \mu_0(1 + \lambda B). \tag{62}$$

The flux density at diameter y is

$$B = \mu H = \frac{0.4Ni\mu_0}{y}(1 + \lambda B). \tag{63}$$

Solving this equation for B, and integrating from d_i to d_0 gives the total flux in a unit height of core, from which the mean flux density can be calculated. This gives for the mean permeability, approximately

$$\mu_m \doteq \mu_0 \left(1 + \lambda B_m \frac{d^2}{d_i d_0} \right), \tag{64}$$

where B_m is the mean peak flux density in the core, and d is the effective magnetic diameter. Comparison of (64) with (62) shows that the increase of mean permeability in a core of considerable radial thickness is not precisely equal to the ideal increase of permeability for a given flux density.

The effect of radial thickness of core on losses can be attacked in a similar manner. Inserting the value of H at diameter y in Cauer's

expression for power loss,[21] computing the loss in a ring of thickness $dy/2$, and integrating from d_i to d_0 yields an expression for core resistance identical with (33), (34) and (35) except that the terms containing α or λ must be multiplied by the factor $d^2/d_i d_0$. It appears that radial thickness of the core affects only that part of the loss which depends upon the variation of permeability with H or B.

In preparing a test core, a compromise must be struck between the radial thickness, diameter, and axial height, so as to secure the desired cross-sectional area without excessive length of copper winding. The laminations of the core must be well insulated from each other. They should be of uniform thickness, which should be known accurately. A good technique used with material in the form of ribbon or tape, consists of winding it tightly in several layers upon a cylindrical mandrel and providing insulation against eddy current straying by dusting the strip with finely powdered alumina while winding. Such insulation is found to withstand the high temperatures ordinarily used in heat treating the core. In order to eliminate the airgap in this type of core, the inside end of the tape may be brought out, folded over, and welded to the outside end, before annealing the core. The use of 50 turns of tape or more in a spiral core reduces the airgap effect to a negligible amount so that welding the tape ends is not necessary. A mandrel diameter should be selected large enough to make the ratio $\Delta d/d_m$ quite small. Thus, a 9 cm. mandrel, wound to a depth of 1 cm., gives a core in which the magnetizing force decreases about 20 per cent from the inside diameter to the outside, while the correction term decreases the effective diameter about 0.4 per cent below the mean diameter. The correction to the permeability variation term is $d^2/d_i d_0 = 1.002$.

The completed core, if strain sensitive, can be protected from the mechanical stresses of handling and winding by mounting it in a loose fitting toroidal box upon which the test windings are applied. Such spacing helps to decrease distributed capacitance, but even more important is sectionalizing, or bank winding of the coil. Types of insulation and windings depend upon the loss characteristics of the core. In general, lower loss characteristics in the core require higher quality windings, to permit measurements at higher frequencies.

SYMBOLS

a Hysteresis resistance coefficient.

a_1 Hysteresis loop area; $= \frac{1}{2}aB_m^3 = 4\pi W$.

[21] W. Cauer, *Arch. f. Elektrotechnik* **15**, 308 (1925).

A Cross-sectional area of magnetic core; cm.^{2}

α Permeability-magnetizing force coefficient; $= \mu_0\mu_m\lambda$.

B Instantaneous flux density in core; gauss.

B_m Maximum flux density in core subject to alternating magnetizing force.

c Residual resistance coefficient.

C Distributed capacitance of coil winding; farad.

d Effective magnetic diameter of annular core; cm.

e Eddy current resistance coefficient of core.

e_c Eddy current resistance coefficient of coil winding.

f Frequency of alternating current.

G Distributed leakance of coil winding; mho.

h $= (aB_m + c)$.

H Instantaneous magnetizing force in core; œrsted.

H_m Maximum magnetizing force in core subject to alternating magnetizing force.

i_m Maximum of alternating current wave in coil; ampere.

I Effective or r.m.s. current in coil.

L Inductance, due to core and residual air space only; henry.

L_a' Inductance due to residual air space.

L_a Inductance due to coil with air core.

L_m Inductance with current of maximum value i_m due to core only.

L_{fm} Inductance due to core only, with current of maximum value i_m, at frequency f (i.e., subject to magnetic shielding).

$L_{o.s.}$ Total inductance observed at frequency high enough to give increases due to distributed capacitance and leakance of the coil winding.

λ Permeability-flux density coefficient; $= \dfrac{\mu_m - \mu_0}{\mu_0 B_m}$.

μ_0 Initial permeability of core.

μ_m Permeability corresponding to L_m.

μ_{fm} Permeability corresponding to L_{fm}.

p Ratio of eddy current to hysteresis resistance at maximum frequency.

P Total power dissipated in core; watt.

P_e Eddy current power dissipated in core.

P_h Hysteresis power dissipated in core.

q Fraction of core loss tolerated as error due to capacitance, leakance, and copper eddy current loss in winding.

Q Insulation quality factor $= \omega C/G$.

R Resistance due to core and winding only; ohm.

R_c Direct current resistance of copper winding.

R_{ce} Eddy current resistance of copper winding.
R_e Eddy current resistance due to core.
R_h Hysteresis resistance due to core.
R_{fm} Resistance due to core only, corresponding to L_{fm}.
$R_{\text{obs.}}$ Total resistance corresponding to $L_{\text{obs.}}$.
ΔR Resistance due to core and leakance only.
ρ Resistivity of core material; abohm-cm.
ρ_1 Ditto; microhm-cm.
t Thickness of sheet, diameter of wire, or r.m.s. sphere diameter of magnetic material; cm.
θ Eddy current parameter; $= 2\pi t\sqrt{\mu_0 f/\rho}$ for sheet material.
W Hysteresis energy dissipated per cycle per cubic centimeter of core; erg.

The Present Status of Ferromagnetic Theory *

By R. M. BOZORTH

DISCOVERY of the loadstone and some of its magnetic properties is now reputed to be some three thousand years old. During these many years ferromagnetism has resisted very successfully the attack of theorists, and even at the present time theory lags far behind experiment. But advances in theory have been particularly rapid during the last five or ten years; the author describes in this paper what he regards as the high points of this progress.

Not until the last quarter of the last century was any considerable work done on magnetic materials. During this period data were gathered rapidly until, just before the close of the century, an excellent book [2] of four hundred pages, containing practically all of the important experimental and theoretical facts, was written by J. A. Ewing, later Sir James Ewing. The shape of the magnetization curves of iron, cobalt, and nickel, the existence of magnetic saturation and the magnetic transformation temperature, the existence and some of the laws of hysteresis, the simpler effects of stress and of magnetostriction, together with the important methods of measurement—all were known then, and silicon steel had just been invented.

Strangely enough, during the next fifteen years there was but little advance in knowledge of magnetic materials, but there were many applications of existing knowledge by engineers to electrical machinery, including those in electrical communication. During this period, also, the Heusler alloys (non-ferrous alloys exhibiting ferromagnetic properties) were invented; and although these served to stimulate those interested in the theoretical aspects of ferromagnetism, still there was little progress.

Beginning between 1915 and 1920 and extending to the present, there has been a rapid development on both the experimental and theoretical sides of ferromagnetism. To illustrate the progress that has been made in the improvement of magnetic materials, Table I has been prepared. The improvements made during the last 20 years have resulted from new methods of purification of the materials, new

* This paper as herein published contains a few revisions and additions to the paper as it appeared in the November 1935 issue of *Electrical Engineering*. It is scheduled for presentation at the A. I. E. E. Winter Convention, New York, N. Y., January 28–31, 1936. A subsequent paper in the same field of endeavor is planned in which entirely new material will be presented.

TABLE I

SOME EXTREMES IN THE PROPERTIES OF MAGNETIC MATERIALS AVAILABLE IN 1915, AND IN 1935

Material	Property	Value 1915	Value 1935
Iron	Maximum permeability	45,000[11]	340,000[12]
	Initial permeability	300	20,000[12]
	Coercive force in œrsteds	0.3[11]	0.03[12]
Iron-nickel [13]	Maximum permeability	2,800[14]	600,000[15]
	Initial permeability	700[16]	12,000[17]
	Coercive force in œrsteds	1.5[14]	0.01[15]
Silicon-iron	Initial permeability	400	2,000[12]
Iron	Hysteresis at $B_M = 100$ gausses, in ergs per cm.^{3} per cycle	20	0.1[12]
Iron-cobalt-nickel "perminvar"	Hysteresis at $B_M = 100$ gausses, in ergs per cm.^{3} per cycle		0.00003[18]
Iron-cobalt	Saturation value in gausses [19]	25,800	25,800
	Permeability at $B = 16{,}000$ gausses	2,100	19,000[18, 12]
Tungsten steel	Coercive force [4] in œrsteds	80	80
New K. S. steel	Coercive force [21] in œrsteds		900

Superior numerals refer to references at end of paper.

compositions (alloys), and new methods of heat treatment. Some of these figures refer only to laboratory specimens, and not to materials available in commercial quantities.

But the chief topic of this paper is the theoretical side of ferromagnetism. How is one to explain the different values of magnetic permeability, ranging from 1 to 600,000 for various materials? Or, to consider first the more fundamental questions, what is the elementary magnetic particle, and why is ferromagnetism associated with so few elements?

ORIGIN OF FERROMAGNETISM

It was suggested by Ampère about one hundred years ago that molecules might behave as magnets because of the electric currents circulating in them. Today, with the advance in knowledge of atomic structure, the origin of ferromagnetism can be discussed in more specific terms. Strangely enough, the spectroscopists have supplied, so to speak, the elementary magnetic particle. It is the spinning electron. In order to explain their extensive observations on spectral lines, they found it necessary to revise the picture of the atom. For some time it has been supposed that an atom was made of a heavy nucleus with a positive charge and of electrons moving in circular

or elliptical orbits around the nucleus. To this picture now must be added the idea that each electron itself is spinning about an axis that passes through its center. Thus, there is circulation of electricity in an atom, both around the nucleus and within each electron—and the latter motion is called the "electron spin" because of its similarity to a spinning ball. Each electron in an atom is then a small gyroscope, possessing a definite magnetic moment on account of its moving electrical charge and a definite angular momentum on account of its moving mass. The ratio of these two quantities is known from various independent lines of reasoning and evidence to possess a particular value. Electrons revolving *in orbits* also exhibit both magnetic moments and angular momenta due to their orbital motions, but for these the ratio is just half what it is for the spinning electron.

The Barnett experiment [22] shows in a very direct way the existence of these magnetic and mechanical moments of the electron and confirms the ratio between them in ferromagnetic materials (Fig. 1). A

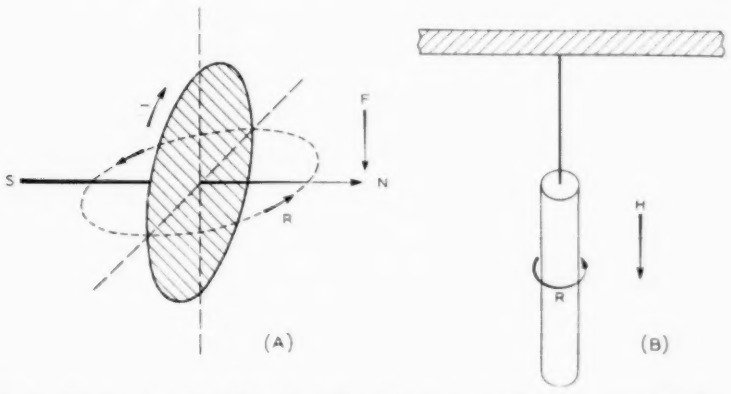
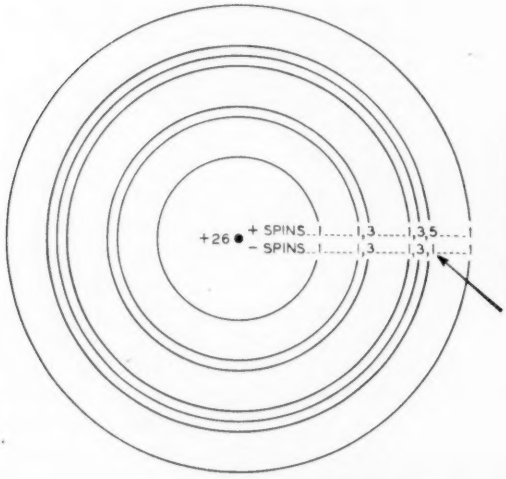


Fig. 1—Gyroscopic action (left); force F produces rotation R. Gyromagnetic effect (right); field H produces rotation R.

rod of iron is hung from a fine suspension and then is magnetized suddenly, whereupon the rod is observed to turn, twisting the suspending fiber a minute but measurable amount. The spinning electrons responsible for ferromagnetism have been turned by the applied field so that they are more nearly parallel to it; but the mechanical moment, which is also a property of those same electrons, causes the whole rod to rotate in just the way that a gyroscope would. Or, to put it differently: When the elementary magnets, pointing originally in all directions, are turned more nearly into parallelism with the axis of

the rod by the applied field, they acquire a net angular momentum parallel to that axis. By the principle that action must be balanced by a corresponding reaction, the rod itself now must recoil with an equal and opposite momentum; it is this last that manifests itself by the sudden twist of the rod and may be calculated from the measured value of the twist. Its sign shows that the spinning magnetic particle is charged negatively, and its magnitude is what would be expected from the hypothesis that that particle is a spinning electron. Thus a change in magnetization is fundamentally a change in the direction of the spin of the electrons in the atom, and not a change in orientation of the whole electron orbit.

The next question is: Why is not every substance ferromagnetic? The picture of the atom of iron as now envisioned by the experts in this field, is represented by the diagram in Fig. 2. The twenty-six elec-



	+ SPINS	– SPINS	EXCESS SPINS
Cr	4	0	4
Mn	5	0	5
Fe	5	1	4
Co	5	2	3
Ni	5	3	2
Cu	5	5	0

Fig. 2—Electron shells in the iron atom.

trons in iron are divided into four principal "shells," each shell a more or less well defined region in which the electrons move in their orbits, and some of these shells are subdivided. The first (inner) shell contains two electrons, the next shell eight, the next fourteen, and the last two. As the periodic system of the elements is built up from the lightest element, hydrogen, the formation of the innermost shells begins first. When completed, the number of electrons in the first four shells are two, eight, eighteen, and thirty-two, counting outward,

but the *maximum* number in each shell is not always reached before the next shell begins to be formed. For example, when formation of the fourth shell begins, the third shell contains only eight electrons instead of eighteen; it is the subsequent building up of this third shell that is intimately connected with ferromagnetism. In this shell some electrons will be spinning in one direction and others in the opposite, and these two senses of the spins may be conveniently referred to as positive and negative. The numbers inserted in Fig. 2 show how many electrons are present in each shell which have positive and negative spins, and it may be noticed that in the iron atom all of the shells except the third contain as many electrons spinning in one direction as in the opposite. The magnetic moments of the electrons in each of these shells mutually compensate one another so that the shell is magnetically neutral and cannot have magnetic polarization. In the third shell, however, which is not yet filled to this extent, there are five electrons with a positive spin and one with a negative so that four electron spins are (unbalanced or) uncompensated and there is a resultant polarization of the atom as a whole. If one more positive charge and its associated mass (a proton) be added to the nucleus and one more electron to an outer shell, the iron is transformed into cobalt; and by repeating the process, the cobalt is transformed into nickel. In iron these additional electrons and their spins are so oriented that there is what may be called an excess spin of four units in iron, three in cobalt, and two in nickel. In manganese, the element just preceding iron in the periodic table, there is an excess of five spins. Only in incomplete shells such as this, shells that are being filled as new and heavier atoms are made, is there such excess spin. The completed shells are magnetically neutral because the spins mutually compensate one another.

The outermost electrons are those responsible for the ordinary chemical properties, and they are influenced by chemical combination. They do not contribute to ferromagnetism for reasons that will appear later.

Exchange Forces

Only in certain parts of the periodic table are there found electrons being added to *inner* shells, and one of these places is in the iron group; but since there are other parts, notably those occupied by the palladium, platinum, and rare earth metals, where these inner groups are being filled, there arises the further question: Why are not these other elements also ferromagnetic?

For an element to be ferromagnetic, it is necessary not only that there be uncompensated spin in the electron orbits, but also that the

resultant spins in neighboring atoms be parallel. Calculation of the energies of the electrons indicates that to align the spins in all the atoms in a small region, the diameter of an atom must bear the proper ratio to the diameter of the electron shell in which the electron spins are uncompensated [23] (Fig. 3). This proper ratio is required because

In ferromagnetic substances, D/d is greater than 1.5 (Slater)

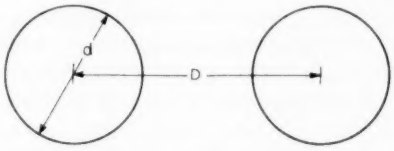


Fig. 3—"Incomplete" shells in neighboring atoms.

the electron spins and charges influence each other to an amount depending upon the distance between them; and it is only when this influence, which is known technically as the "exchange," has the right value that the spins all can be aligned in the same direction,[24] that is, that the material can become ferromagnetic.[25]

The forces of "exchange," the existence of which has been realized only in the last few years, act to keep the spins parallel, while thermal agitation tends, obviously, to disturb this alignment. When the temperature is high enough, the temperature agitation prevails and the material ceases to be ferromagnetic. This temperature is the familiar Curie point, or magnetic transformation point, 770 degrees centigrade (or 1,043 degrees absolute) for iron. It is seen then that the height of the Curie point, θ (on the absolute temperature scale) is an indication of the strength of the forces of exchange, which cannot yet be calculated theoretically except as to order of magnitude. These Curie points are plotted in Fig. 4 for the elements near iron in the periodic table; if a continuous curve be drawn through the points, it has a maximum near cobalt. Now the saturation value of magnetization depends both on the exchange and on the number of effective electron spins, that is, upon the number of electrons that can be oriented parallel to the field and the strength of the forces that hold them parallel. In a very rough way, it may be said to depend on the product of the exchange and the number, S, of uncompensated spins in the atom. Adopting θ as a measure of the exchange forces and forming the product θS, the right-hand curve in Fig. 4 is obtained, which indicates that the highest saturation should be attained in an iron-cobalt alloy, and that under certain appropriate conditions manganese might be ferromagnetic. Both these indications are substantiated by the data: The only known alloys having a higher

saturation value than pure iron are the iron-cobalt alloys; and compounds and alloys of manganese are more magnetic than any others that do not contain iron, cobalt, or nickel. The Heusler alloys, composed of manganese, aluminum, and copper, have a saturation almost as high as nickel, and numerous compounds of manganese are ferromagnetic in a less degree.

The forces of exchange are purely electrostatic in origin. But they are not electrostatic in the classical sense of the word; they are the result of electric charges distributed in space in a definite way. It does not seem to be possible to describe them easily in words, for it takes a great many mathematical equations to derive the result, which is a consequence of the assumptions of quantum mechanics. These forces

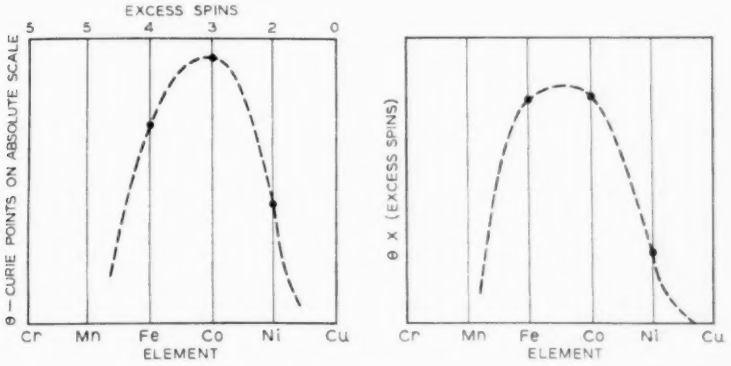


Fig. 4—Iron-cobalt alloys have the highest value of saturation magnetization.

account for the fact that it is easy to align the excess spins of large groups of atoms of some materials. In fact, when the forces of exchange are large as they are in ferromagnetic materials, the stable situation is one in which the spins are parallel, even when no magnetic field is applied. But the parallelism under such circumstances does not extend over the whole of a specimen of ordinary or even of visible size; for some reason not understood it is limited to smaller regions. On the average, those regions are found experimentally to have the volume of a cube about 0.001 inch on an edge. *An actual ferromagnetic body is composed of a great many such regions, called "domains," each domain being magnetized to saturation (i.e., electron spins parallel) in some direction.* When the material is said to be unmagnetized, the domains are oriented equally in all directions so that the magnetization of the specimen as a whole is zero.

Experimental evidence of the existence of these domains is supplied by the so-called "Barkhausen effect" (Fig. 5). If a small portion of a magnetization curve such as is shown in Fig. 5 could be magnified a billion times, it would be seen to be made up of steps, each a sudden change in magnetization as the field is increased, with no further change until the field reaches a certain higher value. No known apparatus can give such direct magnification, but these sudden jumps can be detected by winding a coil around the specimen and connecting its ends to an amplifier at the output of which is a pair of telephone

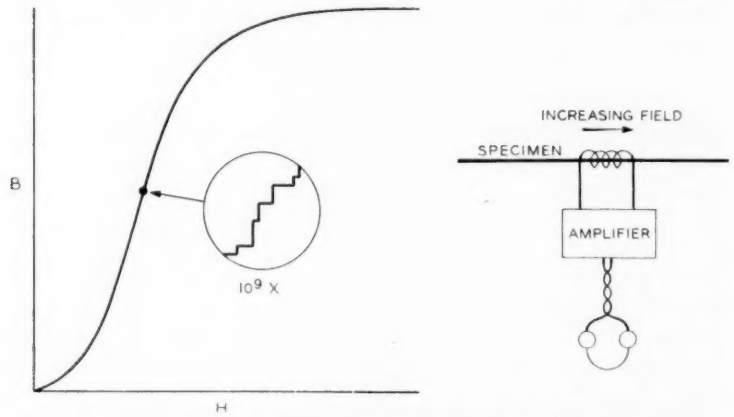


Fig. 5—Sudden changes in magnetization cause the Barkhausen effect.

receivers. When the field is slowly increased, a series of clicks, or "noise," is heard in the receivers; a more quantitative method shows that the average click corresponds to the reversal of magnetization in a region the size [26] of a cube 0.001 inch on an edge, containing 10^{15} atoms. Under favorable conditions this "Barkhausen noise" can be heard without an amplifier, with the receivers connected directly to the coil.

It has been pointed out that the forces of exchange are opposed by the disordering forces of temperature agitation. As a result, the saturation value of magnetization decreases continuously as the temperature is increased, until at the Curie point the ferromagnetism disappears. Data for saturation at various temperatures are shown in Fig. 6, plotted in such units that the saturation is unity at the absolute zero of temperature, and the Curie point is unity on the temperature axis. On such a plot it is found that the data for iron, cobalt, and nickel fall close together. The lower curve is the theoretical one calculated

thirty years ago on the assumption that the elementary magnets, when they are disturbed by temperature agitation, can assume any orientation. If it be assumed, on the contrary, that the spinning electrons responsible for ferromagnetism can assume only two orientations with respect to the other electrons in the atom, the upper curve is the result. If we assume that four orientations are possible the calculated curve lies close to the upper curve of Fig. 6, but somewhat

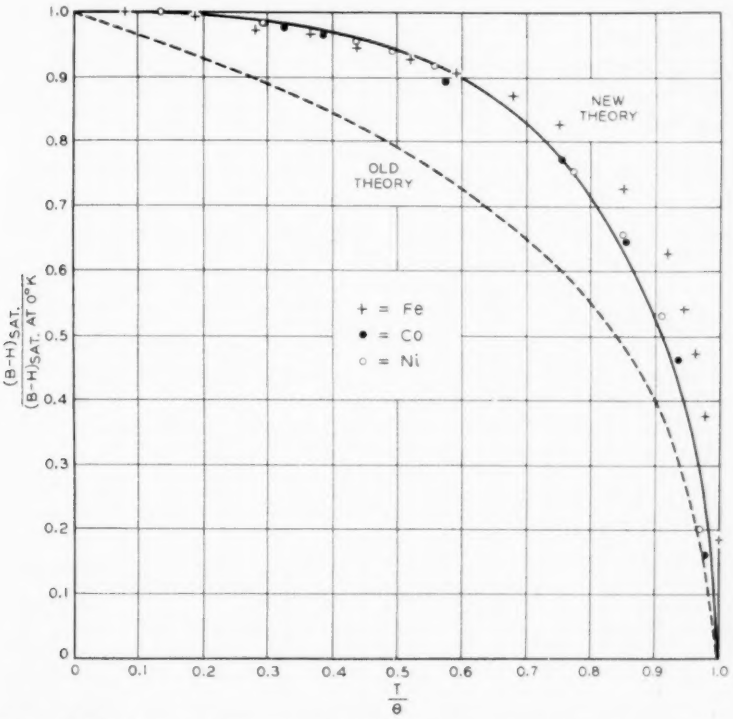


Fig. 6—Dependence on temperature of the saturation magnetization of iron, cobalt, and nickel.

below it, and as the number of possible orientations is increased the curve approaches the lower one shown in the same figure. The close agreement between the data and the upper calculated curve is of special interest because the spectroscopists and atomic structure experts have come independently, each group from its own data, to the conclusion that each electron in an atom can assume only a small number of orientations with respect to the rest of the atom.

Effect of Crystal Structure

There is another kind of force that must be postulated in order to explain the properties of a single crystal. Because of the spinning electrons which it contains, and also because of their orbital motions, each atom may be regarded as a small magnet. These magnets will influence each other in a purely magnetic way,[27] just as a group of bar magnets will; and in a crystal it may be readily appreciated that because of these magnetic forces between atoms arranged in a regular fashion, some directions of magnetization are more stable than others. In iron the most stable direction is observed to be that of the cube edge, one of the cubic axes of the crystal. In nickel it is the cube diagonal (Fig. 7).

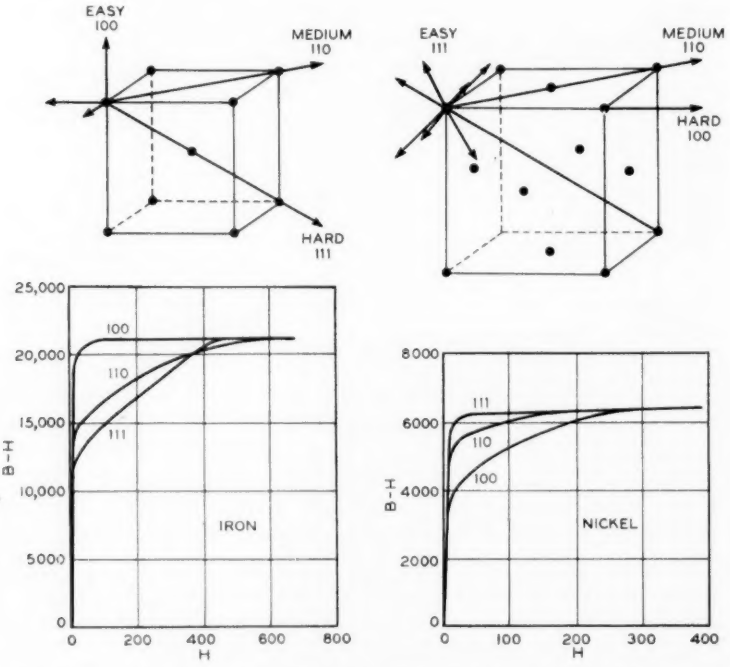


Fig. 7—Magnetic properties and crystal structures of single crystals of iron and nickel (Beck, Honda and Kaya, Webster).

Ordinarily a piece of iron is composed of crystal grains each one of which is too small to be detected by the naked eye. In recent years, however, means have been found to control the grain size of all the

common metals, and single grains (i.e., single crystals) have been prepared which are so large that experiments may be performed and data collected on just one such crystal.

The structure of a single crystal of iron may be represented by a cube with an atom at each corner and one in the center, the whole crystal made up of such cubes packed together face to face. It is found experimentally that in the direction of an edge of this cube (called by the crystallographers a [100] direction) the magnetization curve labeled 100 in Fig. 7 is obtained.[28] In the two other principal directions, the direction of a face diagonal and that of a cube diagonal, the other magnetization curves are obtained, as shown. The difference in the initial parts of the magnetization curves is negligible, the effects being large only above half saturation.

The structure of nickel may be represented also by an assemblage of cubes, but the atoms are arranged in a different manner, being at the corners of the cubes and the centers of the cube faces (Fig. 7). The magnetization curves for nickel corresponding to the same three principal directions are shown also in Fig. 7, and it may be seen that the curves are reversed in order from those of iron. In iron the [100] direction is said to be the direction of easy magnetization and the [111] the direction of most difficult magnetization, whereas the reverse is true in nickel. It might be said that the electrostatic exchange forces align the spins parallel to each other and that the crystal forces determine the particular crystal direction along which they shall be aligned. The forces of exchange are so powerful that they are able to align the spins of a group of atoms, a situation that in the absence of such exchange forces could be accomplished at room temperature only by an applied field of 10,000,000 œrsteds. On the other hand, the crystal forces are so feeble that it takes only 1,000 œrsteds to redirect the spins of an entire group of atoms from any direction to any other direction. The ratio between these two equivalent fields is thus 10^7 divided by 10^3, or 10^4.

As a result of the forces of exchange and the magnetic crystal forces in a single crystal of iron, for example, the situation is as represented in Fig. 8. Even when the crystal is apparently unmagnetized, or demagnetized, there are small regions, called domains, that are magnetized to saturation in one of the six equivalent directions of the crystal axes. Actually, the domains vary considerably in size and shape, but are represented conveniently as squares. Each of the six directions is equally stable and equally probable when no field is applied. The initial effect of applying a magnetic field is to change the direction of magnetization from one stable position to another, thus

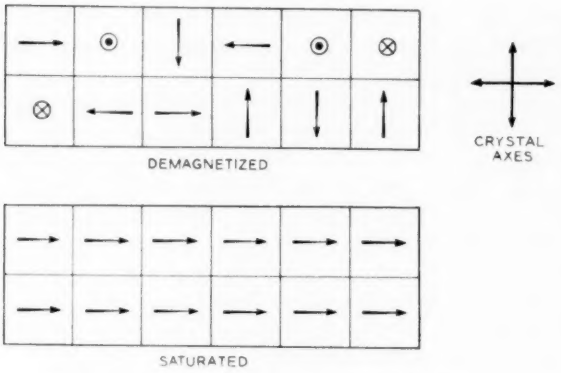
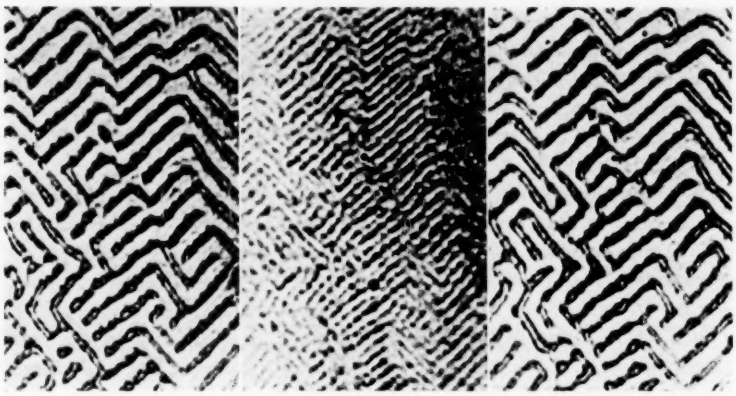


Fig. 8—Domains in a single crystal of iron.

increasing the resultant magnetization in the direction of the field. These changes take place suddenly, and they are the cause of the Barkhausen effect; each sudden change in orientation of a domain accounts for one step in the magnified magnetization curve shown in Fig. 5, or for one click heard in the telephone receiver when listening to the Barkhausen effect.

There is even more direct evidence of the existence of domains in a piece of iron. The iron is placed under a microscope with a magnification of 500, and is covered with a colloidal suspension of iron oxide.[29] It is found (Fig. 9) that the colloidal particles are concentrated along



(a) (b) (c)

Fig. 9—Powder-patterns for iron (McKeehan and Ellmore): (left) field outward; (middle) demagnetized; (right) field inward.

lines determined by the crystal axes, indicating that stray magnetic fields go in and out of the surface just as if some sections were magnetized differently from their neighbors. This occurs even when the iron is unmagnetized, but never occurs with materials that are not ferromagnetic.

Now consider in more detail by what processes changes in magnetization occur. Most changes are attributable to the reorientation of electron spins in domains, from one direction of easy magnetization to another (Fig. 10). These are the changes that take place over the

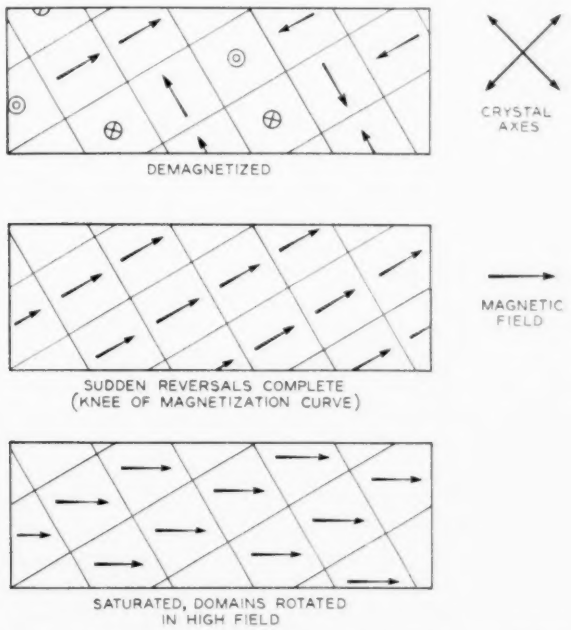


Fig. 10—As the magnetic field strength increases, domains first change direction suddenly, then rotate smoothly.

large central portion of the magnetization curve. In general, however, it is obvious that this process is complete before the material is saturated. When all the domains are magnetized parallel to that direction of easy magnetization which is nearest to the direction of the applied field, the only way in which the magnetization can be increased further is by rotating the electron spins in each domain out of the stable position toward the field direction. Such a process is described loosely as the "rotation of the domain." This is the process that

occurs in high fields, of the order of 10 to 100 œrsteds; as may be seen in Fig. 7, its beginning corresponds to the place where the curves suddenly bend over, away from the almost vertical section. It is only when the field is applied to a single crystal in the direction of easiest magnetization that this last process is avoided. When the field is applied in the direction of most difficult magnetization, the rotational process begins at a field-strength lower than in any other case.

One other important property of single crystals is accounted for by this picture. This property is evident when a field is applied to a single crystal in a direction not parallel to a principal axis. For example, let the field be applied 30 degrees from a cubic axis of an iron crystal, as indicated in Fig. 11 by the longest arrow. As this field is

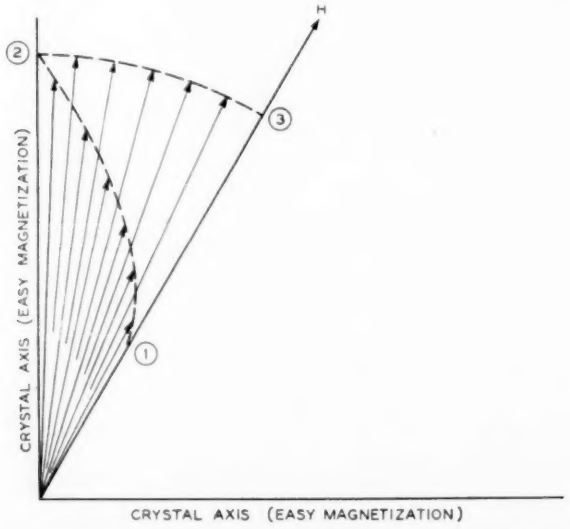


Fig. 11—Vectors represent B-H in iron, increasing in magnitude as the magnetic field (H) increases. First B-H is parallel to H (1); then as B-H increases it deviates in direction from H (2); and finally in high fields is again parallel to H (3).

increased from zero, the magnetization will correspond in magnitude and direction to the other arrows shown. First it is parallel to the field, but as the field increases it deviates toward a direction of easy magnetization until finally it is saturated in that direction. As the field is increased further, the magnetization approaches again the direction of the field and finally is saturated in this direction. Theory

agrees with experiment in that it predicts [30] the direction and amount of the deviation of B from H for any given value of B.

The two ways of changing magnetization that have been described for single crystals, namely sudden changes to new directions of easy magnetization, and continuous rotation of domains, apply equally well to ordinary polycrystalline material, the properties of the latter being those of the former averaged for all orientations. One result of this averaging, of course, is that the specimen is now isotropic and B is parallel to H.

These last remarks must be qualified, for the magnetic materials used by engineers are not always isotropic, that is, the crystal axes are not always distributed equally in all directions. It has been known for many years that when a metal sheet is rolled, the crystals composing it tend to be oriented in special ways with respect to the direction of rolling and to the rolling plane. Even after the sheet has been annealed and recrystallized, these special orientations exist, in some metals all the way up to the melting point. Since the magnetic properties depend on the crystal direction in a single crystal, it follows that sheets composed of crystals having special orientations will not have the same magnetic properties in all directions. This was observed some years ago in iron, nickel, and iron-nickel alloys.[31] More recently, there has appeared on the market a silicon-iron [32] alloy for which the permeabilities in different directions are markedly different. Measured parallel to the direction of rolling this material has a permeability in high fields (B = 15,000 gausses) of 4,000, while measured at right angles to the direction of rolling the permeability is only 400. X-ray analysis shows [33] that the crystals in this material are aligned so that most of them have a cubic axis lying within a few degrees of the direction of rolling. Thus the direction of rolling coincides with the direction of easy magnetization.

In considering the properties of single crystals, the properties in very low fields have not been considered, chiefly because precise data for single crystals are very difficult to obtain. The process that occurs in this region in single crystals and polycrystals must be different from either of the two so far considered, because in ordinary polycrystalline material there are no discontinuities in magnetization, i.e., no Barkhausen effect, and also the fields are not strong enough to rotate the domains to any significant extent against the crystal forces, out of a direction of easy magnetization. Knowing the relation between magnetic force and angular displacement in high fields, it is calculated that if this same mechanism applied to changes in magnetization in very low fields the highest value of initial permeability in iron would

be about 20 instead of many thousands. In the past the process occurring in low fields has been the cause of much speculation, but recently a satisfactory explanation seems to have been found.[5] The changes that take place here are visualized as displacements of the boundaries of domains (Fig. 12); the transition region of a few atom

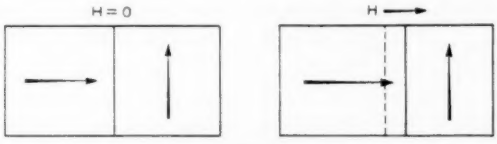


Fig. 12—Magnetization in very low fields progresses by slight displacement of domain boundaries (Becker).

diameters (calculated from the forces of exchange to be about 30 atom-diameters [7]) moves so as to enlarge a domain magnetized in the direction of the field at the expense of a domain pointing in a less favorable direction. Such a movement can progress for only a short distance compared with the linear dimensions of a domain, and is limited by the strains present in any actual material.

Thus in the magnetization of an ordinary well-annealed ferromagnetic material three processes occur, corresponding to the three well known sections of the magnetization curve (Fig. 13): growth of one

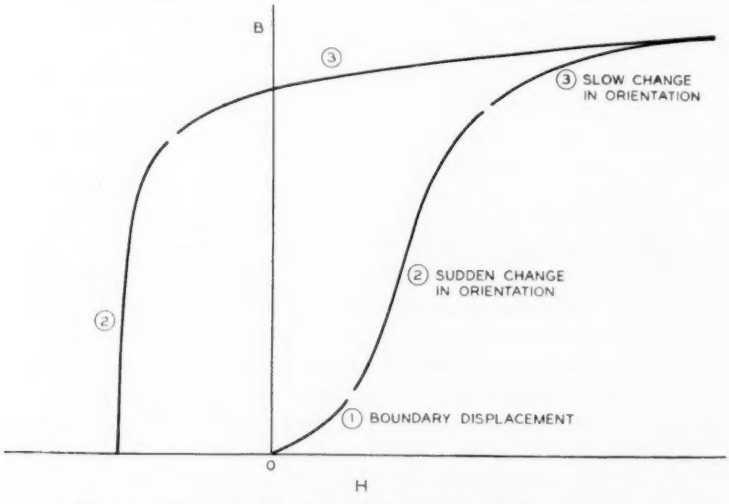


Fig. 13—Illustrating the three kinds of change in magnetization.

domain at the expense of a neighboring one in the initial portion of the curve, sudden changes of direction of domains (with resulting large energy losses) in the middle portion, and continuous or smooth rotation of the domains in the upper portion. The latter two processes occur during the traversal of a large hysteresis loop with tips at high flux densities; the first process is important only in low fields after demagnetization.

Effect of Strain

This picture of the changes in magnetization has been made for materials that are free from any considerable strain. As a matter of fact, strain can affect magnetization in an important way, and under certain circumstances a tensile stress of 5,000 pounds per square inch may change the flux density B as much as 10,000 gausses [34]—almost from zero magnetization to saturation (Fig. 14). The effect is illustrated well by data for 65 and 85 permalloy (iron-nickel alloys containing, respectively, 65 and 85 per cent nickel). For 65 permalloy the effect of tension is to increase the magnetization in all fields; for 85 permalloy the effect is the opposite; and in each case the effect of compression is opposite to that of tension. For ordinary iron the effect of tension is to increase the magnetization in small fields, but to decrease it in high fields.

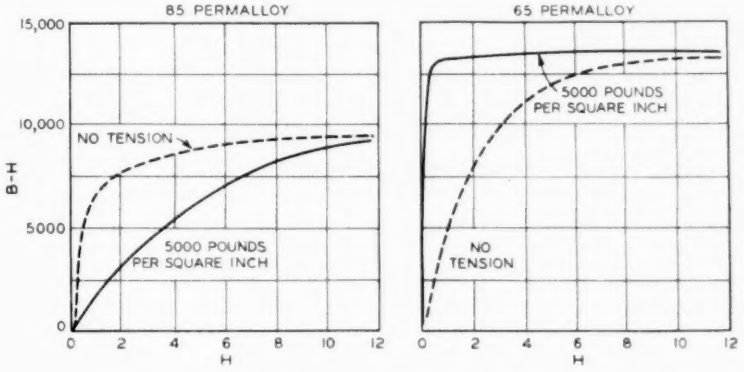


Fig. 14—Effect of tension on magnetization (Buckley and McKeehan).

The effect of strain on magnetization has its counterpart in an effect of magnetization on the length of a piece of ferromagnetic material. When a rod of iron is magnetized its length increases by a small amount. This is but one example of a large class of effects exhibited

by all ferromagnetic bodies and known collectively as "magnetostriction." Figure 15 shows the data for change in length of rods of nickel, iron, and two alloys, plotted against the field H on the one hand and against relative B-H on the other. When saturation of magnetization is reached, the limiting value of magnetostriction, called "saturation magnetostriction," also is attained. Its values for some

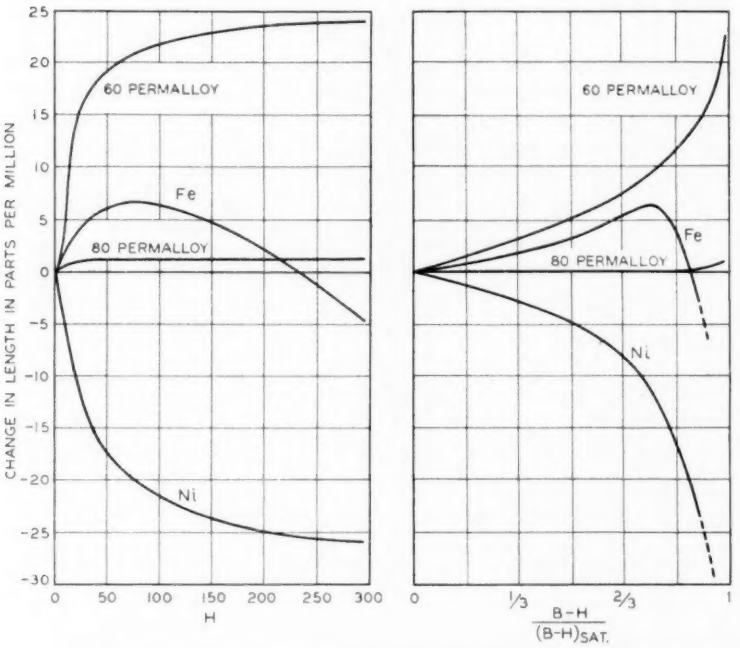


Fig. 15—Magnetostriction in iron, nickel, and two iron-nickel alloys (permalloys).

iron-nickel alloys are shown in Fig. 16. Note here that the change in length is an extension in the alloys containing less than 81 per cent nickel, a contraction otherwise. There is a close relation between magnetostriction and the effect of strain on magnetization, it being a general rule that when the magnetostriction is positive (increase in length with magnetization) the effect of tension is to increase magnetization, and vice versa (Figs. 14, 15, and 16).

How much can theory say of magnetostriction and the effect of strain on magnetic properties? Figure 17 shows how the atoms are arranged in an iron crystal; each atom here is supposed to have a

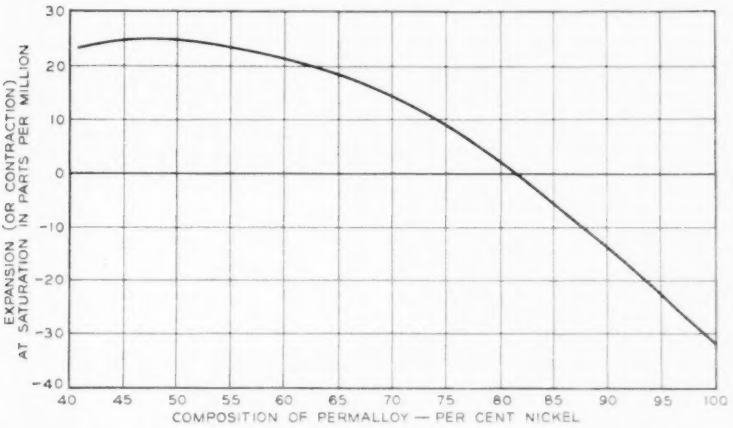


Fig. 16—Saturation magnetostriction of the permalloys (McKeehan and Cioffi, Schulze).

definite magnetic moment as a result of the spin and orbital motion of the electrons. This supposition makes it possible to calculate the magnitude of the mutual magnetic forces which are opposed by the elastic forces holding the crystal together. For iron, the calculations [5] indicate that equilibrium is reached when there has been a slight increase in length in the direction of magnetization and a decrease in length at right angles to this direction such that the volume remains practically unchanged. This calculated magnetostriction is in agreement with experiment as to sign and order of magnitude. With nickel the agreement is not so satisfactory. But in each case the theory is clear in predicting the proper qualitative relationship between magnetostriction and change in magnetization caused by strain.

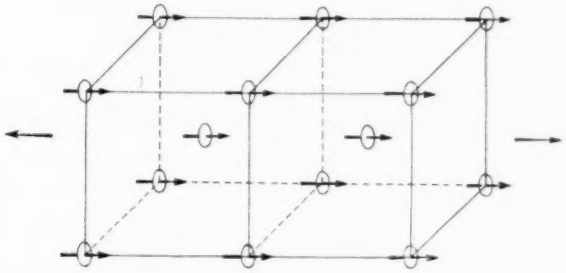


Fig. 17—The magnetic forces between atoms cause a slight elongation in iron (magnetostriction).

Thus magnetostriction and the magnetic effects of strain are reciprocal properties, and result from the same kind of magnetic forces between atoms as those that account for the variation in magnetic properties in different directions in a crystal. Just as the crystal structure determines a direction of easy magnetization in a strain-free crystal, so the strain controls the direction of easy magnetization when the strain is sufficiently great. Figure 18 shows how the domains are

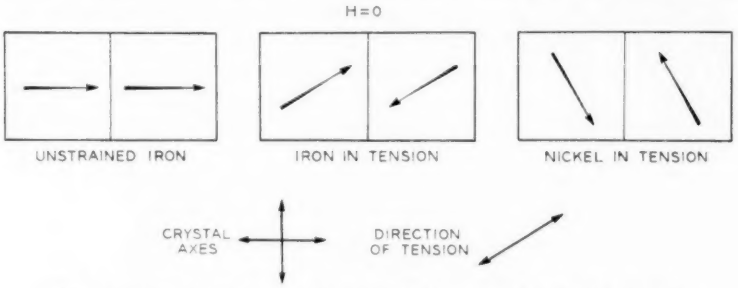


Fig. 18—Domains are oriented by crystal forces and by strain; $H = 0$.

magnetized parallel to the crystal axes in unstrained iron, and how a sufficiently large tension will orient the magnetization parallel to the direction of tension in iron and at right angles to the direction of tension in nickel. When the stress is as large as 10,000 to 30,000 pounds per square inch, the strain effect begins to predominate over the crystal effect and the direction of magnetization is determined mainly by the strain. The calculations show also that in a material having positive magnetostriction the magnetization is increased by tension. In a qualitative way these considerations explain the increase in permeability of 65 permalloy (having positive magnetostriction) and the decrease in 85 permalloy (with negative magnetostriction). But so far the theory is quite inadequate to predict the magnitude of the effect.

In addition to uniaxial homogeneous strains, such as those produced by stretching a wire in the direction of its length, random (heterogeneous) strains are often found that vary in magnitude, sign, and direction from point to point throughout a material. Such strains are produced by cold working, phase transformations, and the like. In such materials the direction of magnetization in a domain is determined by the local strain, and is more stable the larger the strain. So it can be appreciated that it is harder to change the magnetization of a material that is more severely hard worked. These internal strains are the same ones that contribute to the hardness of a metal—hence the parallelism between magnetic hardness and mechanical hardness, which is so well known.

This relation between internal strain and permeability is illustrated by the data [35] shown in Fig. 19. The permeabilities of a series of specimens of 70 permalloy tape, originally cold rolled, increase as the annealing temperature is raised. X-ray data (the angular width of the reflected X-ray beam) on these same specimens indicate the magnitude of the internal strains existing, and show that they become progressively less as the annealing temperature is increased, the most rapid change taking place in each case between 400 and 600 degrees centi-

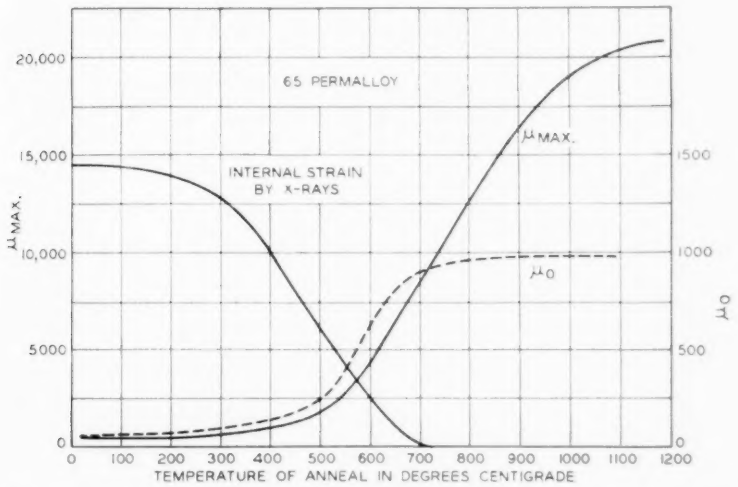


Fig. 19—Magnetic permeability rises as internal strain is diminished by annealing (Dillinger and Haworth).

grade, in which region the microscope shows recrystallization has occurred.

Following out this same idea, it may be surmised that to make good material for a permanent magnet something with very intense internal strains is required. The direct determination by X-rays of internal strain in a good permanent magnet, confirms this view (Fig. 20). Here the widths of the reflected X-ray beams directly measure the internal strains. For comparison with the permanent magnet material, curves are shown for other materials with less internal strain. The magnet material in this case was an iron-nickel-aluminum alloy that was precipitation-hardened, a method used more and more extensively during the last three or four years for such materials. This method is often applicable when the alloy [36] in the stable condition consists of two phases at room temperature (Fig. 21), but when at a

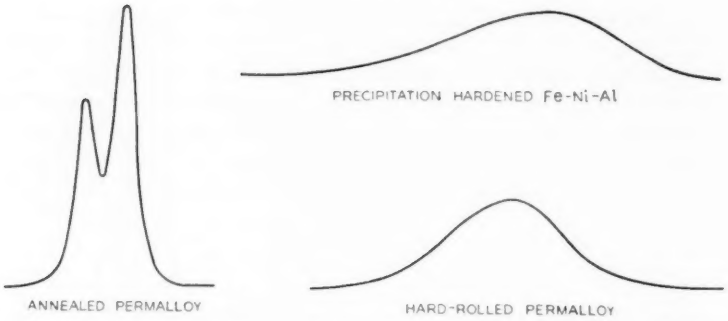


Fig. 20—The width of X-ray reflections indicates the amount of internal strain. Ordinates, intensity of X-rays reflected from metal surface; abscissas, angle of reflection.

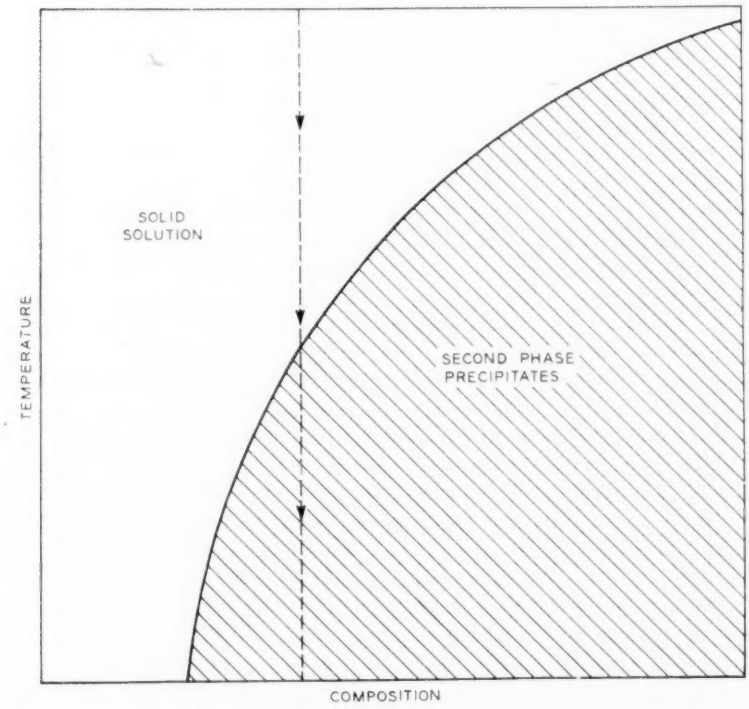


Fig. 21—Precipitation hardening of an alloy for a permanent magnet, such as an alloy of iron, nickel, and aluminum.

higher temperature the one phase dissolves completely in the other to form a solid solution. In making the material, it is quenched rapidly from a high temperature and then reheated to 700 degrees centigrade, at which point the second phase precipitates slowly in very finely divided form. When the optimum amount has precipitated, the material is cooled to room temperature, no more changes occurring. Each submicroscopic precipitated particle is a center of strain, and it is the presence of these unusually large internal strains that is responsible for the good quality of the permanent magnet.

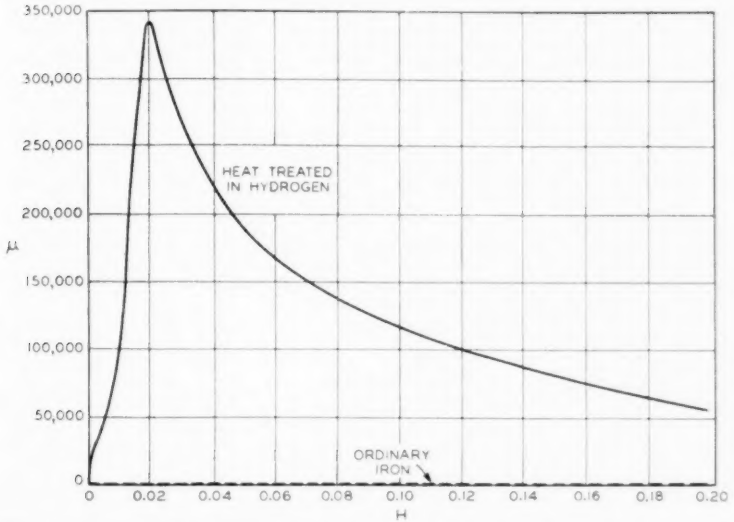


Fig. 22—Permeability curves of ordinary iron and of iron purified by heat treatment in hydrogen at 1,500 degrees centigrade (Cioffi).

Going now to the other extreme, where ease of magnetization is required, it is known, of course, that thorough annealing and a homogeneous structure are beneficial. Still there are at least two sorts of strains that annealing will not relieve. One is that attributable to the non-metallic chemical impurities that do not fit into the regular arrangement of atoms in a pure metal or alloy. It has been found recently that by heat treating iron in hydrogen at about 1,500 degrees centigrade the non-metallic impurities are largely removed, and that what are called "chemical strains" are much reduced. As a result it is found (Fig. 22) that the maximum permeability increases from 10,000 to 340,000,[12] and a large reduction in mechanical hardness occurs simultaneously.

After the chemical strains and the strains resulting from cold working have been removed, there is still another kind of residual strain—that attributable to magnetostriction. These are ordinarily random in direction because they are associated with randomly oriented domains, but by a suitable trick they all can be oriented so as to favor magnetization in a single desired direction at the expense of ease of magnetization at right angles. This trick is heat treatment in the presence of a magnetic field. Without going into a more detailed explanation, the experimental results obtained [15] about two years ago will be given.

When an annealed specimen of 65 permalloy is heated for a few minutes at 650 degrees centigrade while it is subjected to a magnetic field of 10 œrsteds, the maximum permeability is increased from about 20,000 to over 600,000 as shown in Fig. 23. This material holds the

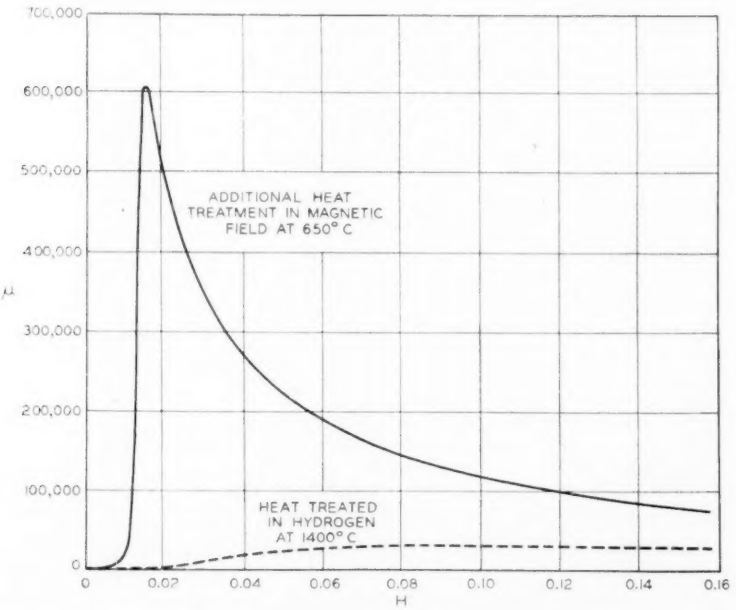


Fig. 23—Permeability curves of 65 permalloy after heat treatment in hydrogen and additional heat treatment in a magnetic field (Dillinger and Bozorth).

records for the highest maximum permeability, the lowest coercive force, and the lowest hysteresis loss at high flux densities. It may be compared with the most permeable material known in 1900, iron with a maximum permeability of less than 3,000.

So far only the effects of stress on the orientation of domains in medium and high fields have been considered. But stress has an effect on the initial permeability also. It has been said already that in very weak fields a change in magnetization is attributed to a movement of the boundaries between domains, the domains oriented nearly parallel to the field growing at the expense of adjacent domains oriented in less favorable directions. Such a growth obviously may be hindered by strain. A relation has been derived [37] connecting the initial permeability with the internal stress and other magnetic quantities:

$$\mu_0 = \frac{0.018(B - H)^2_{\text{sat.}}}{(\Delta l/l)_{\text{sat.}}\,\sigma_i},$$

where μ_0 is the initial permeability, $(B - H)_{\text{sat.}}$ and $(\Delta l/l)_{\text{sat.}}$ are the (ferric) induction and magnetostriction at saturation, and σ_i is the average value of the internal stress in dynes per square centimeter.

Even when there are no internal strains caused by impurities, insufficient annealing, etc., there generally will be the strains of magnetostriction itself, and these will hinder the growth of one domain at the expense of another (Fig. 24). In this case the stress in the foregoing

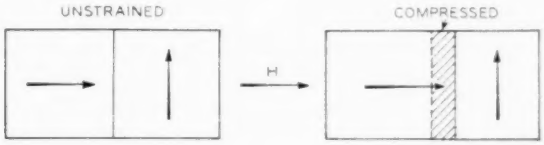


Fig. 24—Magnetostriction in the shaded region acts as a barrier to further change in magnetization.

equation is equal to Young's modulus, E, multiplied by the magnetostrictive strain,

$$\sigma_i = E(\Delta l/l)_{\text{sat.}}$$

and the former equation becomes

$$\mu_0 = \frac{0.018(B - H)^2_{\text{sat.}}}{(\Delta l/l)^2_{\text{sat.}}\,E},$$

This equation really gives a theoretical upper limit for μ_0. These theoretical limits and the highest observed values for iron-nickel alloys are shown in Fig. 25. This indicates why the composition of the "permalloy" having the highest initial permeability is very nearly the same as that for which the magnetostriction is zero.

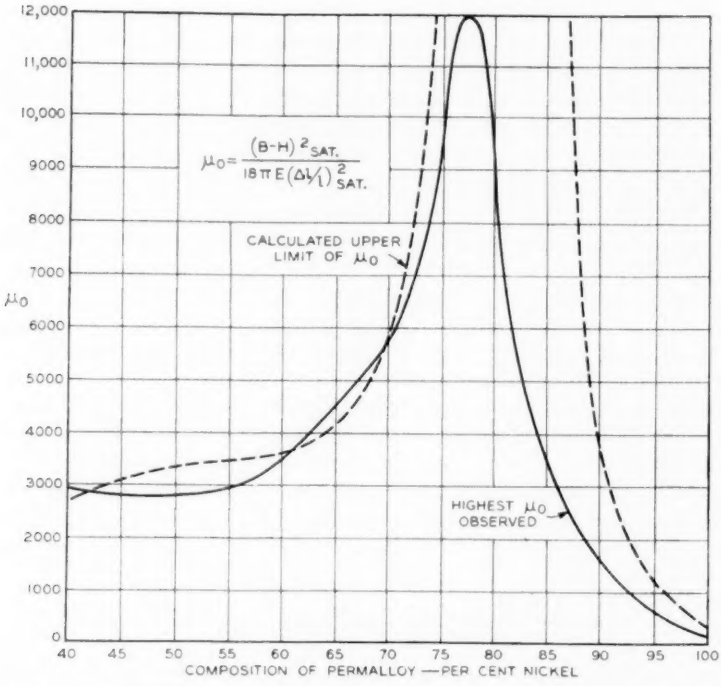


Fig. 25—Comparison of the theoretical upper limit of initial permeability (Kersten) with the highest initial permeabilities observed for iron-nickel alloys (Arnold and Elmen, Schulze).

The effects of strain will now be summarized briefly. The origin of the effects lies in the *magnetic* action between neighboring atoms. The magnetic action is balanced by the elastic (electrostatic) forces between atoms. The balance of these forces results in a change in shape of the magnetic body when it is magnetized (magnetostriction), and also a change in magnetization resulting from strain (strain-sensitivity). Magnetization may be either aided or hindered by a homogeneous uniaxial strain, the effect depending on the magnetostriction in a way that can be estimated qualitatively but not quantitatively. But material in which local strains are directed at random is more difficult to magnetize because the strains prevent a change in magnetization; and the more intense such strains are, the harder the material is to magnetize or demagnetize. The effect of local strains upon the initial permeability can be calculated with fair success, but other magnetic quantities, such as maximum permeability, can as yet be estimated in a qualitative way only.

Summary

In concluding the author wishes to go back from here to summarize what is known about the origin of the forces responsible for the various magnetic properties and about the sizes of the various units. This information is summarized in Table II.

TABLE II

Summary of Data Regarding Origin of Forces Responsible for Various Magnetic Properties

Unit Concerned	Property	Origin of Property	Size of Magnetic Unit
Electron	Magnetic moment	Electron spin	One unit of spin per electron
Paramagnetic atom	Magnetic moment	Uncompensated spins and orbital motions of electrons	4, 3, and 2 uncompensated spins per atom in Iron, Cobalt and Nickel, respectively
Domain	Ferromagnetism. Change in properties at Curie point	"Exchange" between electrons in neighboring atoms	Volume of domain is about $(0.001 \text{ inch})^3$
Single crystal or region of homogeneous strain	Crystal anisotropy. Magnetostriction. Strain sensitivity	Magnetic forces between atoms	10^8 domains per cubic centimeter
Polycrystal	Orientation-average of single crystals and strain units	Sum of effects of single crystals and strains	Size of specimen

Acknowledgment

I take pleasure in acknowledging the detailed criticisms given by Dr. K. K. Darrow, and also the more general criticism given by Dr. O. E. Buckley during the preparation of this paper.

References

Most of the material used in this article can be found in more technical form in the following books and summarizing articles. For some topics reference must be made to the original works, as indicated in the text.

Books

1. "Magnetism and Matter," E. C. Stoner. Methuen, London, 1934.
2. "Magnetic Induction in Iron and Other Metals," J. A. Ewing. Electrician, London, 3d ed., 1900.
3. "Magnetic Properties of Matter," K. Honda. Syokwabo, Tokyo, 1928.
4. "Die ferromagnetischen Legierungen," W. S. Messkin and A. Kussmann Springer, Berlin, 1932.

Recent Summarizing Articles

5. "Theorie der Elektricität," R. Becker. Teubner, Leipzig, chap. 4, v. 2, 1933.
6. "Magnetisms," R. Becker and R. Landshoff, *Die Physik*, **3**, No. 2, 1935, p. 100–8.
7. "Handbuch der Radiologie," F. Bloch. Akad. Verlag., Leipzig, 2d ed., v. 6, pt. 2, 1934, p. 604.
8. "Handbuch der Physik" (Geiger u. Scheel), H. Bethe. Springer, Berlin, 2d ed., v. 24, pt. 2 1933, p. 585.
9. "Stand der Forschung und Entwicklung auf dem Gebiet der ferromagnetischen Werkstoffe," A. Kussmann. *Archiv. für Elektrotechnik*, **29**, May 1935, p. 297–32.
10. "Magnetic Materials," W. C. Ellis and E. E. Schumacher. *Metals and Alloys*, **5**, Dec. 1934, p. 269–74, and **6**, Jan. 1935, p. 26–9.

Original Sources

11. "The Magnetic Properties of Some Iron Alloys Melted in Vacuo," T. D. Yensen. *A. I. E. E. Trans.*, **34**, 1915, p. 2601–41.
12. "New High Permeabilities in Hydrogen Treated Iron," P. P. Cioffi. *Phys. Rev.*, **45**, May 15, 1934, p. 742; and unpublished data.
13. These are without additions of other elements except for deoxidation. With additions of copper and molybdenum, material with initial permeability as high as 40,000 has been reported by H. Neumann. *Arch. f. tech. Mess.*, **4**, Dec. 1934, T-168.
14. "The Magnetic and Electrical Properties of the Iron-Nickel Alloys," C. F. Burgess and J. Aston. *Met. and Chem. Engg.*, **8**, Jan. 1910, p. 23–6.
15. "Heat Treatment of Magnetic Materials in a Magnetic Field," J. F. Dillinger and R. M. Bozorth. *Physics*, **6**, Sept. 1935, p. 279–91. See also earlier paper "Permeability Changes in Ferromagnetic Materials Heat Treated in Magnetic Fields," G. A. Kelsall. *Physics*, **5**, July 1934, p. 169–72.
16. G. Panebianco, Rend. d. Napoli, **16**, 1910, p. 216.
17. "Permalloy, an Alloy of Remarkable Magnetic Properties," H. D. Arnold and G. W. Elmen. *Franklin Inst. Jour.*, **195**, May 1923, p. 621–32.
18. "Magnetic Properties of Perminvar," G. W. Elmen. *Franklin Inst. Jour.*, **206**, Sept. 1928, p. 317–38.
19. "Saturation Value of the Intensity of Magnetization and the Theory of the Hysteresis Loop," E. H. Williams. *Phys. Rev.*, **6**, Nov. 1915, p. 404–9.
20. "The Iron Cobalt Alloy, Fe_2Co, and Its Magnetic Properties," T. D. Yensen. *Gen. Elec. Rev.*, **19**, Sept. 1915, p. 881–7.
21. "New K. S. Permanent Magnetic" (in English), K. Honda, H. Masumoto, and Y. Shirakawa. *Science Rep., Tohoku Imp. Univ.*, **23**, 1934, p. 365–73.
22. S. J. Barnett, *Rev. Mod. Physics*, **7**, 1934, p. 129.
23. "Atomic Shielding Constants," J. C. Slater. *Phys. Rev.*, **36**, July 1930, p. 57–64. See also ref. 8.
24. "Zur Theorie des Ferromagnetismus," W. Heisenberg. *Zeit. Physik*, **49**, July 1928, p. 619–36.
25. In a recent paper G. Urbain, P. Weiss and F. Trombe, *Compt. Rend.*, **200**, 1935, p. 2132, report that the rare-earth metal gadolinium is ferromagnetic. For this substance in the metallic state D/d is about 3, in accordance with Slater's rule (see Fig. 3). This usually high value of D/d indicates that gadolinium should have a low Curie point, which in fact is observed to be 16 degrees centigrade.
26. "Barkhausen Effect, II.—Determination of the Average Size of the Discontinuities in Magnetization," R. M. Bozorth and J. F. Dillinger. *Phys. Rev.*, **35**, April 1, 1930, p. 733–52.
27. "Zur Theorie der Magnetisierungskurve von Einkristrallen," N. S. Akulov. *Zeit. Physik*, **67**, 1931, p. 794. See also "A Contribution to the Theory of Ferromagnetic Crystals," G. Mahajani. *Phil. Trans. Roy. Soc.*, **228A**, 1929, p. 63–114, and reference 5.
28. Here the ordinates are B-H instead of the more usual B, since the former approaches a limiting "saturation" value. B-H sometimes is referred to as "ferric induction."

29. "Surface Magnetization in Ferromagnetic Crystals," L. W. McKeehan and W. C. Elmore. *Phys. Rev.*, **46**, Aug. 1, 1934, p. 226–8, and private communication. See also the earlier experiments of L. v. Hamos and P. A. Thiessen, "Über die Sichtbarmachung von Bezirken verschiedenen ferromagnetischen Zustandes fester Korper," *Zeit. Phys.*, **75**, 1932, p. 562, and "Experiments on the Nature of Ferromagnetism," F. Bitter. *Phys. Rev.*, **41**, Aug. 15, 1932, p. 507–15.
30. "The Theory of the Ferromagnetic Anisotropy of Single Crystals," R. M. Bozorth. *Phys. Rev.*, **42**, Dec. 15, 1932, p. 882–9.
31. "Anisotropie in magnetischen Werkstoffen," O. Dahl and F. Pfaffenberger. *Zeit. Phys.*, **71**, Aug. 1931, p. 93–105.
32. "Discussion on the Influence of Grain Size on Magnetic Properties," N. P. Goss. *Trans. A. S. Metals*, **22**, Dec. 1934, p. 1133–9.
33. "The Orientation of Crystals in Silicon Iron," R. M. Bozorth, *Trans. A. S. Metals*, **23**, Dec. 1935, p. 1107–11.
34. "Effect of Tension Upon Magnetization and Magnetic Hysteresis in Permalloy," O. E. Buckley and L. W. McKeehan. *Phys. Rev.*, **26**, Aug. 1925, p. 261–73.
35. J. F. Dillinger and F. E. Haworth, unpublished data.
36. "Dauermagnetwerkstoffe auf der Grundlage der Auscheidungshärtung," W. Köster. *Stahl u. Eisen*, **53**, Aug. 17, 1933, p. 849–56.
37. "Über den Einfluss des elastischen Spannungszustandes auf die Grösse der Anfangspermeabilität," M. Kersten. *Zeit. techn. Physik*, **12**, 1931, p. 665–9, based on the earlier work "Zur Theorie der Magnetisierungskurve," R. Becker. *Zeit. Phys.*, **62**, May 1930, p. 253–69.

Some Equivalence Theorems of Electromagnetics and Their Application to Radiation Problems

By S. A. SCHELKUNOFF

After a review of the general aspects of the classical electromagnetic theory several "equivalence" theorems are established and illustrated with a number of examples from the diffraction theory. Then follows a discussion of possible applications of these theorems to radiation problems. The latter part of the paper is devoted to the calculation of the power radiated from an open end of a coaxial pair.

THE usual methods of calculating the power radiated by an electric circuit depend upon a determination of the electromagnetic field from the electric current distribution in the circuit. The best known of these methods consists in integrating the Poynting vector over the surface of an infinite sphere surrounding the circuit. This method has been used exclusively until recent years; to facilitate its application, John R. Carson obtained a compact general formula for the radiated power.[1] Another method[2] consists in calculating the work done against the forces of the field in supporting a given current distribution in the circuit. Theoretically either of the two methods is sufficient for solving any radiation problem. Practically, aside from inherent difficulties involved in the calculation of the electric current distribution in the first place, the preliminary integration for determining the field components E and H may be rather complex. Thus in obtaining the power radiated by a semi-infinite pair of perfectly conducting coaxial cylinders this preliminary integration has to be extended over the infinite surfaces of the two conductors. And yet by the Maxwell-Poynting theory, no energy can flow through the walls of the outer cylinders since the electric intensity E and hence the Poynting vector vanish there. Any energy which is radiated away must pass through the open end and it is natural to expect that there must be a method for calculating this energy from the conditions at the open end. The integration involved in this method would extend only over a comparatively small area of the open end. It is in search of a method of this type for calculating the radiated power that I was led some time ago to certain "equivalence theorems." Subsequently I learned that

[1] John R. Carson, "Electromagnetic Theory and the Foundations of Electric Circuit Theory," *The Bell System Technical Journal*, pp. 1–17, January 1927.

[2] A. A. Pistolkors, "The Radiation Resistance of Beam Antennas," *Proc. I. R. E.*, Vol. 17, No. 3 (1929). R. E. Bechmann, "On the Calculation of Radiation Resistance of Antennas and Antenna Combinations," *Proc. I. R. E.*, Vol. 19, p. 1471 (1931).

one of these theorems was discovered long ago, first by A. E. H. Love [3] and then by H. M. MacDonald [4] and proved by the latter [5] for the case of non-dissipative media in 1911. Another proof of this theorem, believed to be helpful from the physical point of view and extended so as to include the dissipative media, is given in this paper. After a brief review of some fundamental concepts we shall prove these equivalence theorems, discuss their significance, and solve one or two simple examples for illustrative purposes.

The physical sources of electromagnetic fields are electric and magnetic charges in motion, that is electric and magnetic currents. The radio engineer has never been interested in shaking magnets for the purpose of radiating energy and has settled into a habit of ignoring magnetic currents altogether as if they were non-existent. It is true that there are no magnetic conductors and no magnetic conduction currents in the same sense as there are electric conductors and electric conduction currents but magnetic convection currents are just as real as electric convection currents, although the former exist only in doublets of oppositely directed currents since magnetic charges themselves are observable only in doublets. And, of course, the magnetic displacement current, defined as the time-rate of change of the magnetic flux, is exactly on the same footing as the electric displacement current defined by Maxwell as the time-rate of change of the electric displacement. We shall find it convenient, at least for analytical purposes, to employ the concept of magnetic current on a par with the concept of electric current.

The two fundamental electromagnetic laws can now be stated in a symmetric form. Ampère's law as amended by Maxwell is: *An electric current is surrounded by a magnetic field of force; the induced magnetomotive force in a closed curve is equal to the electric current passing through any surface bounded by the curve.* In its original form, the "electric current" meant only the conduction current so that the law was applicable only to closed conduction currents. Maxwell's amendment consisted in including the displacement currents, thereby making the law applicable to open conduction currents. The second law is due to Faraday: *A magnetic current is surrounded by an electric field of force; the induced electromotive force in a closed curve is equal to the negative of the magnetic current passing through any surface bounded by the curve.* The rule for algebraic signs is as follows: choose some direction of the closed curve as positive and have an observer placed in such a way that

[3] A. E. H. Love, "The Integration of the Equations of Propagation of Electric Waves," *Phil. Trans.* A, Vol. 197, pp. 1–45 (1901).

[4] H. M. MacDonald, "Electric Waves," p. 16 (1902).

[5] H. M. MacDonald, "The Integration of the Equations of Propagation of Electric Waves," *Proc. London Mathematical Society*, Series E, Vol. 10, pp. 91–95 (1911).

this direction appears to him counterclockwise; then the positive direction of either the electric or the magnetic current is chosen *toward* the observer. If the currents are flowing toward the reader, the directions of the E.M.F. and the M.M.F. are as indicated in Fig. 1.

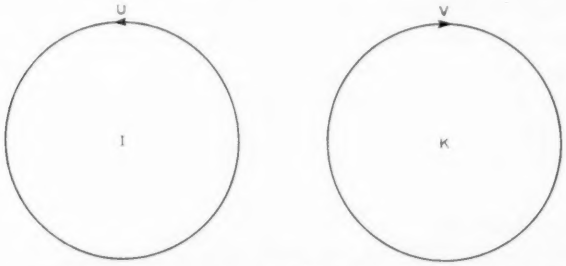


Fig. 1—The relative directions of the E.M.F. and the M.M.F. induced respectively by the electric current I and the magnetic current K are indicated by the arrows. Both I and K are directed toward the reader.

In the well-known way these two physical laws lead to a pair of partial differential equations

$$\text{curl } E = -M, \qquad \text{curl } H = J, \tag{1}$$

where J and M are respectively the total electric current density and the total magnetic current density. The electric density is composed of several parts; namely: the conduction current density, the displacement current density and the applied current density. The first of these components is, in many substances, proportional to the electric intensity E; the second is proportional to $\partial E/\partial t$; and the third is due to forces other than those of the field, mechanical or chemical, for instance. Similarly the magnetic current density is the sum of the magnetic displacement density proportional to $\partial H/\partial t$ and the impressed magnetic current density. Thus, we write

$$\text{curl } E = -M_0 - \mu\frac{\partial H}{\partial t}, \qquad \text{curl } H = J_0 + gE + \epsilon\frac{\partial E}{\partial t}, \tag{2}$$

where J_0 and M_0 are the densities of the impressed currents and the constants of proportionality g, ϵ and μ are respectively the conductivity, the dielectric constant and the permeability.[6]

The functions J_0 and M_0 are supposed to be known functions of coordinates and of time, representing the distribution of the physical

[6] A consistent practical system of units is used in this paper. Thus the E.M.F. is measured in volts, the electric current in amperes, E in volts per centimeter, H in amperes per centimeter, etc. The permeability of vacuum is then $4\pi 10^{-9}$ henries per centimeter and the dielectric constant $(1/36\pi)10^{-11}$ farads per centimeter.

sources in the space-time. If they are zero everywhere and at all times, the only physically significant solution of (2) must be $E = H = 0$ throughout the entire space and at all times.[7] If there are other solutions of (2), they are extraneous and some rule must be found for excluding them. Such extraneous solutions often find their way into mathematical equations because it is usually impossible to express *all* physical conditions by an equation or a system of equations. Naturally these remarks do not apply to a limited region of space or a finite interval of time. In fact, in many physical problems these "extraneous in the large" solutions of (2) can be advantageously used for expressing the general character of electromagnetic phenomena in a limited region and then obtaining, with the aid of the boundary and the initial conditions, the complete answer. But the philosophy of causality demands the dictum "no sources, no field" when considering the whole space-time. It may seem unnecessary to dwell at length on such obvious matters but they happen to be essential in the subsequent discussion if the arguments are to be taken as positive proofs rather than as plausible justifications.

Equations (2) are linear and the principle of superposition is applicable. This is in accordance with physical intuition which tells us that we can subdivide the impressed currents into elementary cells of volume dv, calculate the field due to a typical element, and obtain the total field by integration. For the typical element (2) becomes

$$\text{curl } E = -\mu \frac{\partial H}{\partial t}, \qquad \text{curl } H = gE + \epsilon \frac{\partial E}{\partial t}, \tag{3}$$

everywhere except in the infinitely small volume occupied by the element. The product of the current density and the volume of the element is called the *moment* of the element.

At times the impressed currents are confined to sheets so thin that their thickness can be disregarded without introducing a serious error in the result. This leads to a hypothetical infinitely thin *current sheet*. We pass from a real current sheet to an ideal one by assuming that the thickness of the former decreases and the current density increases in such a way that their product remains constant. This product is called the linear density of the sheet and it represents the current per unit length perpendicular to the lines of flow. The moment of a current element is now the product of the density of the sheet and the area of the element. Finally if the impressed current is confined to a

[7] We assume that all the electric and magnetic charges were originally in the neutral state, in which case their separation could be effected only through their motion. The argument could be extended so as to include purely static fields that may constitute an integral part of the universe but it is of no particular interest to us.

very thin filament, the moment is the product of the current and the length of the element.

It is the moment of the current element that determines its electromagnetic field. If the medium is non-dissipative, the actual expressions for the field components are obtained in terms of an auxiliary function called by Lorentz the *retarded magnetic vector potential*. For an electric current element of moment $p(t)$ this vector potential at any point P is parallel to the current density and is a function of the distance r from the element to P

$$A = \frac{p\left(t - \frac{r}{c}\right)}{4\pi r}. \tag{4}$$

The quantity c has the dimensions of a velocity and it appears that the action of the source travels outward with this velocity. But there is another solution of (3)

$$A_1 = \frac{p\left(t + \frac{r}{c}\right)}{4\pi r}. \tag{5}$$

One might wonder if this solution appertains in any way to the source; that is not the case, however. If the moment $p(t)$ is identically zero prior to some instant $t = t_0$, the field which can legitimately be attributed to the action of this source is also identically zero for any instant $t < t_0$. But (5) implies a non-vanishing field at distant points; it is as if the effect appeared before the cause. Any other solution is a combination of (4) and (5) and has to be rejected on the same grounds.

In terms of this auxiliary vector potential the field components can be expressed as follows

$$H = \operatorname{curl} A, \qquad \frac{\partial E}{\partial t} = \frac{1}{\epsilon} \operatorname{curl} H, \qquad E = \frac{1}{\epsilon} \int_{-\infty}^{t} \operatorname{curl} H \, dt. \tag{6}$$

If the moment is harmonic of frequency f, we regard it as the real part of $pe^{i\omega t}$. Then the vector potential and the field components are the real parts of the following expressions

$$A = \frac{pe^{-i\beta r}}{4\pi r}, \qquad H = \operatorname{curl} A, \qquad E = -i\omega\mu A + \frac{\operatorname{grad} \operatorname{div} A}{i\omega\epsilon}, \tag{7}$$

where the *phase constant*

$$\beta = \frac{\omega}{c} = \frac{2\pi f}{c} = \frac{2\pi}{\lambda},$$

λ is the wave-length, and the time factor $e^{i\omega t}$ is implied.

If the medium is dissipative, we have

$$A = \frac{pe^{-\sigma r}}{4\pi r}, \qquad H = \operatorname{curl} A. \qquad E = -i\omega\mu A + \frac{\operatorname{grad}\operatorname{div} A}{g + i\omega\epsilon}. \tag{8}$$

The quantity σ is the *intrinsic propagation constant* of the medium and is defined by

$$\sigma = \sqrt{i\omega\mu(g + i\omega\epsilon)}. \tag{9}$$

In this case the action of the source at some point is not only delayed by the time needed for the disturbance to travel the intervening distance but also exponentially attenuated.

If instead of an electric current element, we are dealing with a magnetic element, the field components can be expressed in terms of an *auxiliary electric vector potential*. This vector F is given by

$$F = \frac{Pe^{-\sigma r}}{4\pi r}, \tag{10}$$

where the moment P of the element is the product of the magnetic current density and the volume of the element. The field components are then given by

$$E = -\operatorname{curl} F, \qquad H = -(g + i\omega\epsilon)F + \frac{\operatorname{grad}\operatorname{div} F}{i\omega\mu}. \tag{11}$$

In the periodic case the general mathematical solution for the vector potential of an element is found to be a linear combination of any two of the following functions

$$\frac{e^{-\sigma r}}{r}, \qquad \frac{e^{\sigma r}}{r}, \qquad \frac{\cosh \sigma r}{r}, \qquad \frac{\sinh \sigma r}{r}. \tag{12}$$

All of these except the first become exponentially infinite at an infinite distance from the source and cannot be taken to represent the vector potential of a physical source. The last function is finite in any finite region; conceivably it can represent an electromagnetic field in the finite region free from physical sources. If the medium is non-dissipative it is impossible to exclude any of the solutions given by (12) on the grounds of their behavior at infinity—they all vanish there. But we may regard the non-dissipative case as the limit of the dissipative one and in this way establish a rule for finding the proper unique solution.

In the presence of a current sheet, equations (2) are valid on either side of it but not on it. Let us consider a cross-section of an electric current sheet, perpendicular to the lines of flow, and a curvilinear

rectangle $A'B'B''A''$ with two of its sides parallel to the sheet (Fig. 2). We assume that the current flows toward the reader and that $A'A''$ and $B'B''$ are vanishingly small. Since the M.M.F. around this rectangle is equal to the electric current passing through it and since this M.M.F. is merely the difference between the M.M.F.'s along the sides $A'B'$ and $A''B''$, we obtain

$$H_t' - H_t'' = J_l \tag{13}$$

by simply calculating these quantities per unit length of the rectangle. The tangential components of the magnetic intensity are regarded as

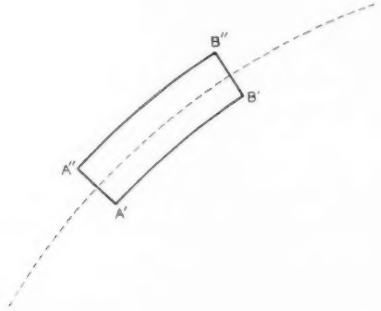


Fig. 2—A cross-section of a current sheet perpendicular to the lines of flow. The positive direction of the current is toward the reader.

positive when directed from A to B. Thus the tangential component of the magnetic intensity is discontinuous across an electric current sheet and the amount of the discontinuity is equal to the density of the sheet.

Similarly across a magnetic current sheet the tangential component of the electric intensity is discontinuous and the amount of this discontinuity is equal to the negative of the magnetic current density of the sheet; thus

$$E_t' - E_t'' = - M_l. \tag{14}$$

In deriving equations (2) it is also necessary to assume that g, μ and ϵ are continuous throughout the region under consideration. They have no meaning on the boundary between two different media. Since the boundary is a geometric surface, it cannot constitute either an electric or a magnetic current sheet. Hence the components of E and H tangential to such a boundary are continuous across it. These *boundary conditions* provide a link between the fields in the two media.

Let us suppose that we have a continuous distribution of sources on a closed surface C (Fig. 3) and that there are no other sources. We assume that the sources are harmonic of frequency f. The electromagnetic field $\mathfrak{F}$ produced by these sources can be calculated directly from this distribution with the aid of the above mentioned vector potentials. On the other hand, we can reason as follows.

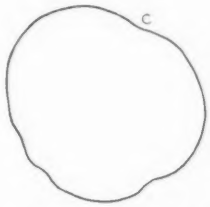


Fig. 3—A cross-section of a closed surface C.

There are no sources either inside or outside of C; hence everywhere except on C, we have

$$\text{curl } E = -i\omega\mu H, \qquad \text{curl } H = (g + i\omega\epsilon)E. \tag{15}$$

In the region inside of C we take that solution of (15) which is finite throughout this region and outside of C we select the solution vanishing at infinity. Both solutions will contain constants which can be determined from conditions (13) and (14) across the surface. The field $\mathfrak{F}'$ obtained in this manner is identical with $\mathfrak{F}$ because the difference $\mathfrak{F} - \mathfrak{F}'$ is everywhere source-free and thus must vanish.

Let us now reverse the process and, instead of starting with the known distribution of sources on C, suppose that we know the field and wish to find its sources. Let the known field $\mathfrak{F}$ be source free everywhere except on C. In order to determine these sources S we merely calculate the discontinuities in the tangential components of E and H across C. We can utilize this result to establish the major *Equivalence Principle*. For the outside portion of $\mathfrak{F}$ we can choose the outside portion of the field $\mathfrak{F}'$ produced by a given system of sources S' situated inside C and for the inside part of $\mathfrak{F}$ we take any field which is source-free there. The latter may be, for instance, the inside portion of the field $\mathfrak{F}''$ produced by some sources S'' situated outside C. Thus we arrive at the following *Equivalence Principle* discovered by Love and Macdonald [8]: a distribution of electric and magnetic currents on a given surface C can be found such that *outside* C it produces the

[8] See references 3 and 4 and also H. M. Macdonald, "Electromagnetism" (1934).

same field as that produced by given sources *inside* C; and also the field *inside* C is the same as that produced by given sources *outside* C. One of these systems of sources can be identically equal to zero.

The actual calculations are made as follows. From the discontinuities in the tangential components of E and H, we obtain J and M by (13) and (14). From these currents we find the two vector potentials

$$A = \frac{1}{4\pi}\int\int_{(C)} \frac{J(x', y', z')}{r} e^{-i\beta r}\, dS,$$
$$F = \frac{1}{4\pi}\int\int_{(C)} \frac{M(x', y', z')}{r} e^{-i\beta r}\, dS, \tag{16}$$

where $r = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}$ is the distance between a point $P(x, y, z)$ somewhere in space and a point $P'(x', y', z')$ on C. From these potentials we calculate the electric intensity and the magnetic intensity by

$$E = -i\omega\mu A + \frac{\text{grad div } A}{i\omega\epsilon} - \text{curl } F,$$
$$H = \text{curl } A + \frac{\text{grad div } F}{i\omega\mu} - i\omega\epsilon F. \tag{17}$$

The proof of the Equivalence Principle can be modified so as to throw some additional light on it. Let us suppose that given sources S' are inside the closed surface C and let us make our new synthetic field by obliterating the old field outside C and leaving everything as it was inside C. The new field has the same sources S' and besides it is discontinuous across C. These discontinuities are the additional sources S whose densities are calculable from (13) and (14). Since the new field is identically zero outside C, the field produced by S is such as to cancel the field produced by S' outside C. Thus the system of sources S acts as a perfect absorber for the electromagnetic wave produced by S'. Reversing the directions of the current distributions on C, we conclude that the system of sources $-S$ produces outside C exactly the same field as S'.

The Equivalence Principle is closely related to another theorem which we may call the Induction Theorem. Let us suppose that a closed surface C subdivides the entire space into two homogeneous media and that a system of sources S is given in one of those regions (Fig. 4). Let E, H be the field due to these sources on the assumption that the medium inside C is the same as that outside. The true field

outside C must vanish at infinity but it need not be the same as E, H; let it be $E + E', H + H'$. The field E', H' must be source-free outside C. Inside C the field must be source-free; we shall designate it by E'', H''. The field E', H' is called the *reflected field* and E'', H'' the *refracted field*. The boundary conditions are such that the components of the electric and the magnetic intensities tangential to C must be continuous. Thus over the surface C, we have

$$\bar{E}_t + \bar{E}_t' = \bar{E}_t'', \qquad \bar{H}_t + \bar{H}_t' = \bar{H}_t''. \tag{18}$$

The bar over the letters is used to designate the values of the corresponding quantities on C. From (18) we obtain

$$\bar{E}_t'' - \bar{E}_t' = \bar{E}_t, \qquad \bar{H}_t'' - \bar{H}_t' = \bar{H}_t. \tag{19}$$

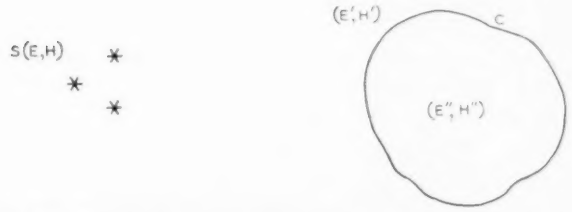


Fig. 4—The closed surface C is the boundary between two homogeneous regions in space. (E, H) designates the field produced by some system of sources S; (E', H') is the field reflected by the body C; and (E'', H'') is the field in the body.

Hence *the reflected and the refracted fields together constitute an electromagnetic field in the entire space; this field is source-free everywhere except on C and the distribution of sources on C is calculable from the given sources S.* This Induction Theorem is a generalization of the well-known theorem used in calculating the response of an electric circuit to an impressed field. Since the wires constituting the circuit are very thin, only the tangential components of E in the direction of the wires need be considered.

It may be noted that if the medium inside C is identical with that outside C, the "reflected field" must be absent and the "refracted field" must be identical with the field E, H due to the sources S. Thus the Induction Theorem leads to the Equivalence Principle.

The Equivalence Principle is evidently an extension of Kirchhoff's theorem. The latter deals with a single wave function instead of two vectors. Kirchhoff derived a formula for computing the wave function in the source-free region from its values and the values of its normal derivative over a closed surface separating the source-free region from

the region containing the sources of the wave functions. In the Theory of Sound the wave function represents the excess pressure or the velocity potential and Kirchhoff's theorem is valuable in the analysis of diffraction phenomena. Kirchhoff's theorem is also used in dealing with optical diffraction. We may also remark that Kirchhoff's formula is a mathematical expression of a principle governing compressional wave motion. This principle was first formulated by Huygens in the following form: each particle in any wave front acts as a new source of disturbance, sending out secondary waves, and these secondary waves combine to form the new wave front.[9]

Let us now examine one of the familiar diffraction problems in the light of the Equivalence Principle. Consider a source S and a perfectly absorbing screen (Fig. 5*a*). Such a screen will be defined in the usual manner: the impressed wave enters it without reflection but does not pass through it. If the screen is infinitely thin, this definition implies the existence of electric and magnetic currents in the screen whose densities are given by the postulated discontinuity in the field. In reality the "black bodies" absorb not by virtue of the coexistence of electric and magnetic currents but by virtue of electric currents alone with the aid of reflections taking place between atomic layers. The true mechanism of absorption is complex and requires more than a mere surface. In diffraction studies it has become a habit with us to ignore the precise nature of absorption and confine outselves to its implications; but it is just as well to know the nature of the ideal mechanism which we are substituting for the true mechanism.

We can apply the Equivalence Principle to the present problem in two ways. We can choose as our surface C a surface (1234) just on the other side of the screen. The part (23) contributes nothing; the equivalent distribution of sources S' over the parts (12) and (34) gives us a complete field to the right of the screen. On the other hand if S'' is the field due to the electric and magnetic currents in the screen induced by S, the total field is $S + S''$. The choice of the "surface C" that would yield this result is shown in Fig. 5*b* although the conclusion is obvious without recourse to the Equivalence Principle. Since to the right of C in Fig. 5*a* the two alternative fields must be the same, we have

$$S' = S + S'' \qquad \text{and} \qquad S' - S'' = S. \tag{20}$$

Incidentally the last equation is the expression for the Equivalence Principle as applied to S in the absence of the screen since $-S''$ is

[9] A. E. Caswell, "An Outline of Physics," p. 544 (1929).

the contribution of that portion of the equivalent layer which was removed by the screen.

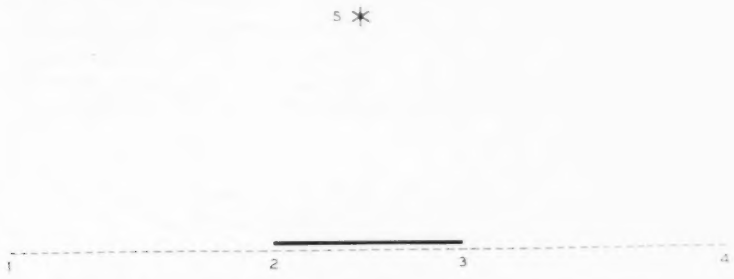


Fig. 5*a*—A source S in front of a screen the cross-section of which is shown in heavy lines.

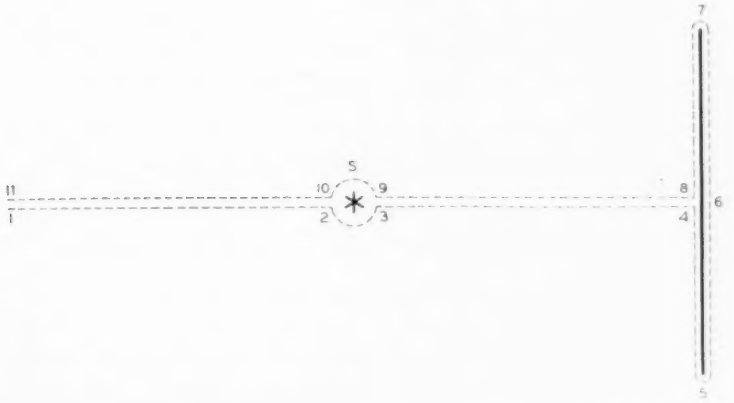


Fig. 5*b*—A source S in front of a screen the cross-section of which is shown in heavy lines.

The case of a hole in a perfectly absorbing screen (Fig. 6*a* and Fig. 6*b*) can be treated in the same manner and the reciprocity existing between this and the preceding case is quite evident. In terms of the sources previously defined the field to the right of the screen is $-S''$; by (20) this is the same as $S - S'$.

If the screen is a perfect conductor, the problem is much more complex. The screen will support electric currents but not magnetic currents. The densities of the electric currents are not calculable directly from the field S but from the condition that the component of the electric intensity tangential to the screen vanishes. The problem

is very difficult and its solution has been found in only a few special cases. It is true that once we know the electric currents in the screen, we can determine the field on both sides of the screen; but there is no simple way of calculating these currents exactly. Frequently it is assumed that, in so far as the side opposite to the source is concerned, a perfectly conducting screen is equivalent to a perfectly absorbing screen of the same geometric character. This is equivalent to a

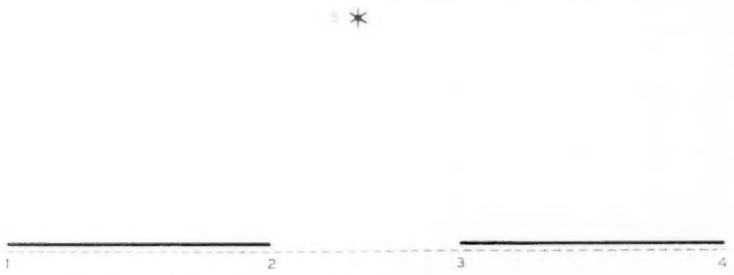


Fig. 6*a*—A source *S* in front of a screen the cross-section of which is shown in heavy lines.

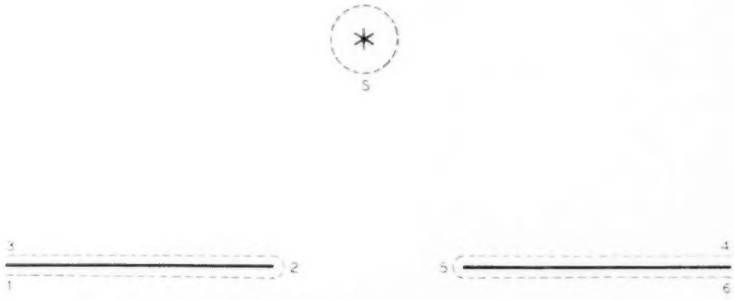


Fig. 6*b*—A source *S* in front of a screen the cross-section of which is shown in heavy lines.

hypothesis that the electric current density of the screen is determined by the magnetic intensity impressed directly by the source *S*. We could take the results obtained from this hypothesis as a first approximation to the true results. The tangential component of the electric intensity calculated on the basis of this hypothesis does not vanish on the screen which means that we have violated the original hypothesis that the screen is a perfect conductor. If the discrepancy is not too great we might look for an additional electric current distribution to reduce this discrepancy.

There are times, however, when the current distribution in the "screen" can be determined with a fair accuracy without elaborate mathematics. It is so, for instance, in the case of a pair of perfectly conducting coaxial cylinders (Fig. 7*a* and Fig. 7*b*) in which the radii

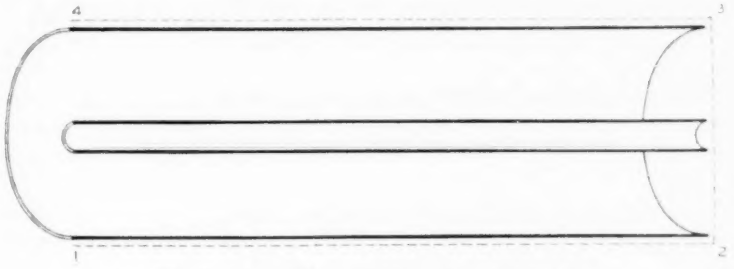


Fig. 7*a*—An axial cross-section of a coaxial pair.

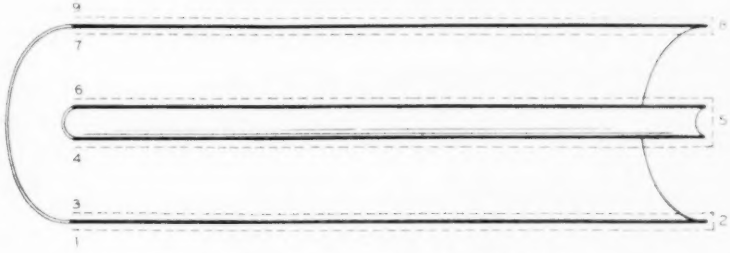


Fig. 7*b*—An axial cross-section of a coaxial pair.

are small by comparison with the wave-length. We shall assume that the coaxial pair is semi-infinite. Trusting his common sense, the engineer assumes that inside this structure the magnetic lines are circles coaxial with the cylinders. The electric lines are the radii and the electric current in the cylinders as well as the transverse voltage between the cylinders vary along the length in the same way as in a transmission line with uniformly distributed series inductance and shunt capacity. A careful analysis by John R. Carson indicates that this simple picture is justifiable if the cross-section of the coaxial pair is small by comparison with the wave-length.[10] While a whole series of electric waves can exist in such a structure, all of these waves except the one recognized by the engineer, the *principal wave*, are attenuated very rapidly and are significant only very close to the generator and

[10] John R. Carson, "The Guided and Radiated Energy in Wire Transmission," *A. I. E. E. Journal*, pp. 908–913, October, 1924.

very near the open end. The complementary waves are needed only for logical consistency and to satisfy the boundary conditions.

Thus let us suppose that the field distribution in the coaxial pair is known to a high degree of accuracy. In order to calculate the field outside the coaxial pair and hence obtain the radiated power we can use the Equivalence Principle in two ways. We can fit our surface C smoothly over the outer cylinder and the open end (Fig. 7*a*) or, regarding this surface as a perfectly elastic rubber sheet, we can press it through the open end and fit it smoothly over the inner surface of the outer conductor and the outer surface of the inner conductor (Fig. 7*b*). Since by hypothesis the conductors are perfect, the components of E tangential to the cylinders vanish; hence in the second choice of C the equivalent layer consists of only an electric current sheet. Naturally this current distribution is precisely that which actually exists in the conductors so that this choice of C leads to something that we knew beforehand, namely: if the actual sources, that is, if the electric currents in the structure are known exactly or approximately, the entire field can be calculated exactly or approximately.

The first choice of C is more important. Over the lateral portion (12, 34) of C the equivalent magnetic current sheet vanishes as in the preceding case on account of the perfect conductivity of the cylinders. The magnetic intensity just outside the coaxial pair is also zero except near the open end where it must be exceedingly small. To see this, we need only recall that the electric currents in the two cylinders are equal and opposite and that except in the neighborhood of the open end the displacement currents are transverse. Thus the equivalent electric current sheet can be ignored altogether. What is left is the magnetic current sheet over the surface of the open end; the density of this sheet is determined by the radial component of the electric intensity and in the final analysis by the voltage existing between the ends of the inner and outer conductors. Presently we shall carry out the actual calculations but just now we shall examine the question of the accuracy of the results. Of course, the results would be exact if we knew the equivalent electric and magnetic sheets accurately; and the above approximations appear to be reasonable. We shall not be able to find out how good these approximations are but we can prove that they are just as good as the approximations usually made in calculating the radiated power from the distribution of electric currents. The only virtue of the Equivalence Principle is to save a certain amount of mathematical work and furnish a further insight into the phenomena of radiation.

If a progressive wave is advancing from left to right in a semi-infinite coaxial pair (Fig. 7) and if the generator is at infinity, we can assume it to be the principal wave. At the open end this wave is reflected. It is usually assumed that the reflected wave is also the principal wave but moving in the opposite direction. In other words, it is assumed that the total field is such that the electric lines are radial and the magnetic lines are circular. Since the electric lines are radial, there is no longitudinal displacement current; and since the conduction current at the open end must be zero, the magnetic intensity is zero over the entire open end. This is what follows if we neglect the complementary waves.

These approximate results correspond to the *exact* results in the following hypothetical situation. If a hypothetical perfect magnetic conductor is fitted over the open end of the coaxial pair so that it closes it entirely, then the reflection is complete and there are no complementary waves. Perfect magnetic conductors are defined by analogy with perfect electric conductors—the tangential component of the magnetic intensity vanishes at the surface of the former just as the tangential component of the electric intensity vanishes at the surface of the latter. Magnetic conductors support magnetic current sheets just as electric conductors support electric current sheets. The densities of the sheets are given by the discontinuities of the tangential components of E in the former case and H in the latter.

In the hypothetical case in which the open end is closed with a perfect magnetic conductor, no energy can flow outside the coaxial pair. This is because the flow of energy is given by $\frac{1}{2}E \times H$ and either one or the other factor vanishes over the outer boundary of the structure. The field outside the coaxial pair must now be identically zero. Our sources are the electric current in the walls of the coaxial pair and the magnetic currents in the cap. If one field is designated by S and the other by S', then $S + S' = 0$ and $S' = -S$. Thus the field produced by the electric currents in the conductors on a supposition that principal waves alone exist, is the same as the field produced by a hypothetical distribution of magnetic currents over the surface of the open end.

Let us examine another case. It is usual to assume that the electric current in a thin wire (Fig. 8) in free space is distributed sinusoidally. Experimental evidence shows that the radiated power calculated on this assumption is very nearly correct. On the other hand the sinusoidal distribution of the electric current in the wire corresponds to a hypothetical case in which a perfect magnetic conductor is introduced

in the shape of a sphere concentric with the center of the wire and passing through its ends. Thus we could calculate the radiated power from an appropriate distribution of magnetic currents over this sphere but in this case such a procedure would involve more difficult integrations than the usual method.

Before considering the more general case of radiation from a semi-infinite coaxial pair let us assume that the radii of the two conductors are nearly equal. We have seen that in applying the Equivalence Principle we need take into account only the magnetic current sheet over the open end of the pair. In the present instance this sheet is merely a circular loop of magnetic current equal to the voltage V between the ends of the conductors. If we were to treat in the same

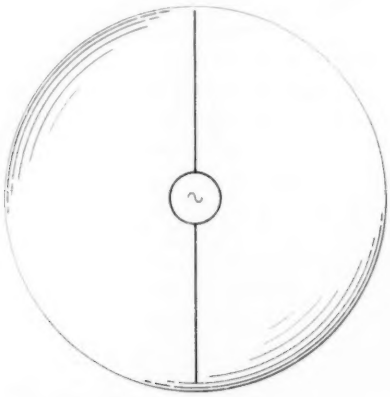


Fig. 8—A vertical antenna and a cross-section of an imaginary sphere passing through the ends of the antenna.

manner a condenser made of two parallel circular plates, we should come to the conclusion that it is also equivalent to a magnetic loop around its periphery. Thus in both cases the radiated power is the same. But the power radiated by an electric doublet is known to be $(40\pi^2 I^2 l^2)/\lambda^2$ watts where I is the amplitude of the electric current, l the length of the doublet and λ the wave-length. In applying this formula to a condenser it is better to express it in terms of the voltage V across the condenser and its area S. The capacity of the condenser is $C = S/(36\pi 10^{11} l)$ farads and $I = \omega CV = SV/60\lambda l$ amperes. Hence the power radiated by the condenser is $(\pi^2 S^2 V^2)/90\lambda^4$ watts. This is also the approximate power radiated by the coaxial pair if we interpret S as the cross-sectional area of either conductor.

Let us now calculate the more general expression for the power radiated from an open end of a coaxial pair. The cylindrical conductors whose cross-sections are shown in Fig. 9 are supposed to extend

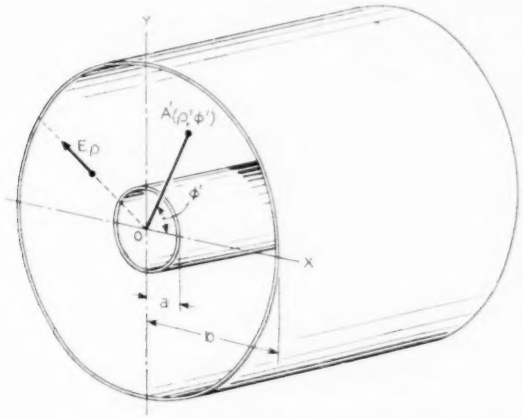


Fig. 9—The end view of a coaxial pair of cylindrical conductors.

below the surface of the paper and the z-axis of the coordinate system is directed toward the reader. The primed letters will refer to points situated in the opening, the unprimed letters being reserved for typical points in space.

The electric intensity in the coaxial pair varies inversely as the distance from the axis

$$E_{\rho'} = \frac{P}{\rho'}, \qquad E_{\varphi'} = 0. \tag{21}$$

In accordance with the Equivalence Principle we assume that the field below the xy-plane is wiped out and the discontinuity in E arising as the result of the separation is replaced by a magnetic current sheet. This magnetic current is perpendicular to E; by (14) its density is

$$M_{\varphi'} = -E_{\rho'} = -\frac{P}{\rho'}, \qquad M_{\rho'} = 0. \tag{22}$$

The constant E is related to the voltage V between the ends of the coaxial conductors; in fact, we have

$$V = \int_a^b E_{\rho'} d\rho' = P \log \frac{b}{a}, \qquad P = \frac{V}{\log \frac{b}{a}}. \tag{23}$$

In order to calculate the field at some point A due to the distribution (20), we must determine the retarded electric vector potential. Since the integration is vectorial, it is convenient to deal with the cartesian components of the magnetic current density

$$M_{x'} = \frac{P \sin \varphi'}{\rho'}, \qquad M_{y'} = -\frac{P \cos \varphi'}{\rho'}. \tag{24}$$

The area of the element is $\rho' d\rho' d\varphi'$ so that the components of the retarded potential are

$$\begin{aligned} F_x &= \frac{1}{4\pi} \int_a^b \int_0^{2\pi} \frac{M_{x'} e^{-i\beta AA'} \rho' d\rho' d\varphi'}{AA'} \\ &= \frac{P}{4\pi} \int_a^b \int_0^{2\pi} \frac{e^{-i\beta AA'} \sin \varphi'}{AA'} d\rho' d\varphi', \\ F_y &= -\frac{P}{4\pi} \int_a^b \int_0^{2\pi} \frac{e^{-i\beta AA'} \cos \varphi'}{AA'} d\rho' d\varphi', \quad \beta = \omega \sqrt{\mu\epsilon} = \frac{\omega}{c} = \frac{2\pi}{\lambda}. \end{aligned} \tag{25}$$

Hence the components in the polar coordinates are

$$\begin{aligned} F_\varphi &= -F_x \sin \varphi + F_y \cos \varphi \\ &= -\frac{P}{4\pi} \int_a^b \int_0^{2\pi} \frac{e^{-i\beta AA'} \cos (\varphi - \varphi')}{AA'} d\rho' d\varphi', \\ F_\rho &= 0. \end{aligned} \tag{26}$$

The distance AA' is

$$AA' = \sqrt{r^2 - 2r\rho' \cos \vartheta + \rho'^2}, \tag{27}$$

where r is the distance OA, and ϑ is the angle between OA and OA'. Since we are interested in the radiation field alone, we need retain only those terms in (24) which vary inversely as the distance; the other terms contribute nothing to the radiated power. Thus we let r increase indefinitely, obtaining

$$AA' = r - \rho' \cos \vartheta, \tag{28}$$

and then

$$F_\varphi = -\frac{Pe^{-i\beta r}}{4\pi r} \int_a^b \int_0^{2\pi} e^{i\beta\rho' \cos \vartheta} \cos (\varphi - \varphi') d\rho' d\varphi'. \tag{29}$$

If θ and θ' are the angles made by OA and OA' with OZ, we have

$$\begin{aligned} \cos \vartheta &= \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos (\varphi - \varphi') \\ &= \sin \theta \cos (\varphi - \varphi'). \end{aligned} \tag{30}$$

Since ρ' is small by comparison with the wave-length λ, we can expand the exponential term in the integrand into a power series and retain

only the first two terms

$$e^{i\beta\rho' \cos \vartheta} \doteq 1 + i\beta\rho' \cos \vartheta = 1 + i\beta\rho' \sin \theta \cos (\varphi - \varphi'). \quad (31)$$

We need the second term because the integral of the first vanishes. Integrating the second term, we obtain

$$\begin{aligned} F_\varphi &= -\frac{i\beta P e^{-i\beta r} \sin \theta}{4\pi r} \int_a^b \rho' d\rho' \int_0^{2\pi} \cos^2 (\varphi - \varphi') d\varphi' \\ &= -\frac{1}{8} i\beta (b^2 - a^2) P \frac{e^{-i\beta r}}{r} \sin \theta. \end{aligned} \quad (32)$$

The magnetic current is uniform around the axis and there is no accumulation of magnetic charge anywhere; hence the second term in the expression for H as given by (11) vanishes. Therefore

$$H_\varphi = -\frac{P}{8} \omega\epsilon\beta (b^2 - a^2) \frac{e^{-i\beta r}}{r} \sin \theta. \quad (33)$$

At a great distance from the source the wave tends to become plane so that in the radiation field the electric intensity is perpendicular to OA and to H and is given by

$$E_\theta = \sqrt{\frac{\mu}{\epsilon}} H_\varphi = 120\pi H_\varphi. \quad (34)$$

According to Poynting the radiated power is the real part of the following integral

$$W = \frac{1}{2} \int_0^\pi \int_0^{2\pi} E_\theta H_\varphi^* r^2 \sin \theta \, d\theta \, d\varphi \text{ watts}, \quad (35)$$

where H_φ^* is the complex number conjugate to H_φ. Substituting from (31) and (32), we obtain

$$\begin{aligned} W &= \frac{\pi^3}{980} \left(\frac{b^2 - a^2}{\lambda^2} \right)^2 P^2 \int_0^\pi \sin^3 \theta \, d\theta \int_0^{2\pi} d\varphi \\ &= \frac{\pi^4}{360} \left(\frac{b^2 - a^2}{\lambda^2} \right)^2 P^2 \text{ watts}. \end{aligned} \quad (36)$$

Introducing from (21) the expression for P and designating by S the area of the opening, we have

$$W = \frac{\pi^2}{360} \left(\frac{S}{\lambda^2 \log \frac{b}{a}} \right)^2 V^2 \text{ watts}. \quad (37)$$

The effect of radiation on the transmission line can be simulated by a resistance R,

$$R = \frac{180}{\pi^2}\left(\frac{\lambda^2 \log \frac{b}{a}}{S}\right)^2 \text{ohms} \tag{38}$$

shunted across the open end. This is not the resistance seen by the generator. If V and I are the amplitudes of the voltage and the electric current at their antinodes and Z_0 the characteristic impedance of the coaxial pair, then

$$V = Z_0 I = \left(60 \log \frac{b}{a}\right) I. \tag{39}$$

Since the end of the coaxial pair is a voltage antinode, the radiated power may be expressed as

$$W = 10\pi^2 \left(\frac{S}{\lambda^2}\right)^2 I^2 \text{ watts.} \tag{40}$$

Hence the radiation resistance seen by a generator placed at a current antinode is

$$R_G = \frac{20\pi^2 S^2}{\lambda^4} \text{ ohms.} \tag{41}$$

With this simple illustration, we conclude the present paper.

Magnetic Alloys of Iron, Nickel, and Cobalt *

By G. W. ELMEN

The unexpected magnetic properties of certain alloys of iron and nickel discovered some 20 years ago led to a thorough study of the entire range of iron-nickel alloys. The results of this study were so encouraging that alloys of these metals with cobalt, the only other ferromagnetic metal, also were studied, as well as various alloys of these metals with small amounts of non-magnetic metals added. From the results of this extended investigation have emerged several alloys that are playing important parts in the continued advancement of electrical communication.

SOME alloys of iron, nickel, and cobalt have remarkable magnetic properties superior in many situations to those of the constituent metals. Many of these alloys have found wide use in the instrumentalities and circuits of electrical communication, and were developed primarily for that purpose. This paper reports the experience and techniques of the Bell Telephone System in the development and utilization of these materials.

The advantageous properties of these alloys were disclosed through exhaustive researches, during which the whole realm of combinations of these three metals was explored. That certain alloys of iron and nickel had unexpected properties at low flux densities had already been discovered in the Bell Telephone Laboratories. There was at that time no theoretical basis for predicting, or even explaining, the character of those alloys; and, therefore, a study was undertaken of the whole iron-nickel series. The results were so encouraging that combinations of these elements with cobalt likewise were studied; and finally those alloys of special interest were combined with varying amounts of non-magnetic metals. In the course of this investigation several thousand specimens were made and tested in a period extending over fifteen years.

Such an empirical investigation is time consuming and expensive, but in a field where so little theory was available for guidance it was the only certain means to determine the practical possibilities of these alloys. It has been justified by the large number of alloys it has developed for practical use in communication engineering. One of the first and most striking applications was to submarine telegraph cables. The largest field of application, however, has been in teleph-

* Published in the December 1935 issue of *Electrical Engineering*, and scheduled for presentation at the Winter Convention of the A. I. E. E., New York City, January 28–31, 1936.

ony, where the requirements generally are very exacting, and where other advances have imposed rigid demands on the magnetic materials.

In telephone circuits, standards of transmission efficiency require that the magnetic materials used as circuit elements shall produce maximum magnetic effect with minimum energy loss and distortion of the transmitted currents. Translated into magnetic characteristics, this means that at low magnetizing forces the material shall have high permeability in combination with low hysteresis loss, and, in many situations, constancy of permeability over the operating range. In circuits for voice and carrier currents it is often necessary to reduce the intrinsic permeability of the material to obtain the required constancy and low losses in the apparatus. Furthermore, to minimize eddy currents, a high resistivity is required and the material must be structurally suitable for fabrication into thin laminations. For other uses, such as for signaling and switching mechanisms, the magnetic properties at medium and high flux densities determine the suitability of the material. High permeability and low coercive force make for improved sensitivity and speed of operation. Low coercive force is of special interest in marginal apparatus where the difference between the operating and releasing currents is small.

Preparation and Composition of the Alloys

A great many factors contribute to the final properties of an alloy. Among the most important of these are the purity of the elements used in the alloy, their preparation, and the heat-treatment. The magnetic properties attainable can be completely masked by the intrusion of small quantities of certain impurities or by improper heat-treatment. For iron the magnetic properties can be improved materially by removing extremely small quantities of carbon and other non-metallic elements through heat-treating * in an atmosphere of hydrogen and at temperatures close to the melting point. This method of purification also improves the magnetic properties for alloys of iron, nickel, and cobalt. For communication purposes, it has not been found expedient as yet to introduce this method of refinement in the commercial production of these alloys. The purity of the constituents is controlled by ordinary methods of chemical analysis, by methods of melting, and by annealing processes which do not increase the amounts of important

* There is a rapidly growing technical literature relating to the effects of very small percentages of impurities on magnetic properties and the methods for their removal, with notable contributions by T. D. Yensen of the Westinghouse Electric and Manufacturing Company, W. E. Ruder of the General Electric Company, and P. P. Cioffi of the Bell Telephone Laboratories.

impurities. The magnetic properties recorded in this paper have therefore been confined mostly to those obtained on materials produced by standard metallurgical methods.

In the commercial method of producing these alloys the best grades of commercial iron, nickel, and cobalt are used. The melting is done in an electric furnace, and after the mechanical fabrication into suitable shapes these alloys are heat-treated to develop the desired magnetic properties.

Early in an investigation of these alloys it was found that some of them required special heat-treatments to develop the desired magnetic properties. For some the slow cooling incident to the ordinary process of annealing was not suitable, and a rapid cooling was necessary. For another group the slow cooling in the annealing process was not slow enough, and the best results were obtained when the alloys were held at a constant high temperature for a considerable time. It was evident that to determine the most suitable temperature of heating and rate of cooling for each alloy would require more time than was warranted in the exploratory work. Three methods of heat-treatment that, in a general way, would separate the alloys into groups, were developed. These heat-treatments are designated in this paper as "annealing," "quenching," and "baking."

The annealing process consists of heating the samples in closed containers to a temperature of 1,000 degrees centigrade, and cooling with the furnace. The cooling ordinarily requires 7 hours before room temperature is reached. This heat-treatment is primarily for the purpose of removing the effects of mechanical strains necessarily resulting from the rolling and stamping of the alloys into suitable shapes. All the alloys discussed in this paper received this heat-treatment before any of the more special processes were applied.

The quenching process consists of heating the alloys for a short time at 600 degrees centigrade, and cooling in air at room temperature for small samples with large surfaces, and in oil for larger samples. The rate of cooling attained by these methods is approximately 40 degrees centigrade per second. It has been found that the best rate of cooling for maximum permeability does not always develop the highest initial permeability. The difference, however, is not large, and often is masked by other variations in the manufacturing process.

The baking process consists of heating the alloys for 24 hours at 425 degrees centigrade, and then slowly cooling to room temperature. The rate of cooling does not affect the development of the magnetic properties unless it is so rapid as to introduce mechanical strains.

Classification of the Alloys

A convenient way of showing graphically the compositions of the alloys of iron, nickel, and cobalt is by means of the composition triangle in Fig. 1. The sides of this triangle represent the binary alloys of the three metals, and points inside the triangle, the ternary alloys.

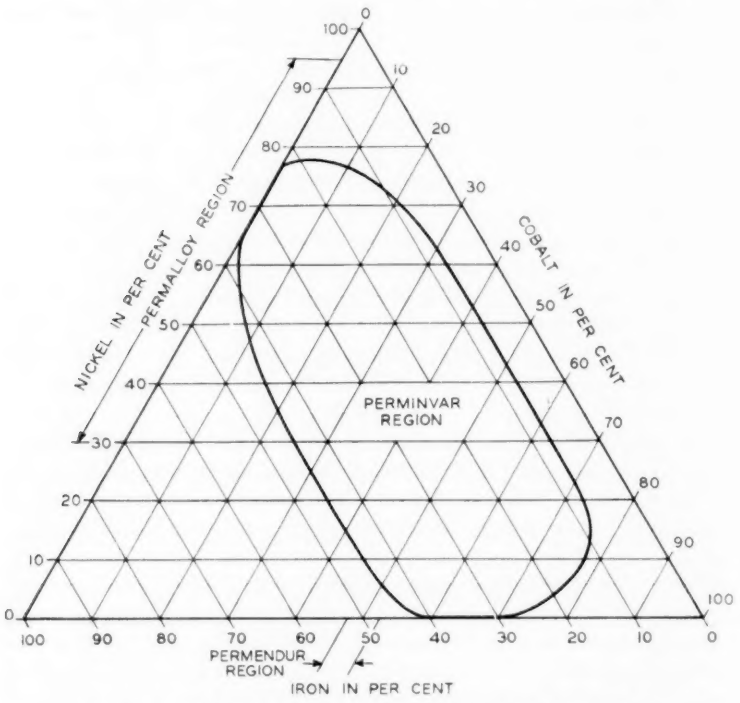


Fig. 1—Composition diagram for alloys of iron, nickel and cobalt.

In this diagram the alloys of special interest because of their magnetic properties are indicated, and, for convenience, each group in which the magnetic properties are similar has been given a specific name.

On the iron-nickel side of the triangle the permalloy region is indicated. In this group several compositions have been developed for commercial use in the Bell System. The method of identifying these alloys consists of prefixing a numeral indicating the per cent of nickel; for example, 45-permalloy contains 45 per cent nickel and 55 per cent iron. To some of these permalloys small amounts of other metals also

are added. In designating ternary permalloys the same scheme is extended, so that the name gives everything except the iron content, and this is obtainable by difference. Thus, 3.8–78.5 Cr-permalloy contains 3.8 per cent chromium, 78.5 per cent nickel, and 17.7 per cent iron.

The perminvar region, enclosed by the curved line, contains those compositions that require baking to develop completely their characteristic magnetic properties. The specific compositions of these alloys are indicated by two prefixed numerals, the first indicating the nickel and the second the cobalt percentages, respectively. Thus the 45–25 perminvar alloy contains 45 per cent nickel, 25 per cent cobalt, and 30 per cent iron. Another alloy of the perminvar group, in which the nickel and cobalt percentages are the same as the alloy just mentioned, but which contains 7 per cent molybdenum and 23 per cent iron, is designated as 7–45–25 Mo-perminvar.

In the iron-cobalt series of alloys the composition 50 per cent iron and 50 per cent cobalt has been developed for commercial use. This is the permendur alloy, indicated in the triangular diagram in Fig. 1. This alloy is difficult to cold roll, but the addition of 1.7 per cent vanadium improves the mechanical properties and makes it sufficiently ductile to roll into thin sheets. The same system has been followed in designating this alloy as in the case of the permalloys. Thus 1.7 V-permendur is an alloy containing 1.7 per cent vanadium with iron and cobalt in equal proportions.

Table I lists the designations and compositions of those alloys, developed for particular purposes, which are discussed more fully in the remainder of this paper.

TABLE I

DESIGNATIONS AND COMPOSITIONS OF SOME MAGNETIC ALLOYS

Designation	Composition, Per Cent					
	Ni	Fe	Co	Cr	Mo	V
78.5 permalloy	78.5	21.5				
80 permalloy	80	20				
45 permalloy	45	55				
3.8–78.5 Cr-permalloy	78.5	17.7		3.8		
3.8–78.5 Mo-permalloy	78.5	17.7			3.8	
2–80 Mo-permalloy	80	18			2	
45–25 perminvar	45	30	25			
7–45–25 Mo-perminvar	45	23	25		7	
Permendur		50	50			
1.7 V-permendur		49.15	49.15			1.7

Ni = nickel; Fe = iron; Co = cobalt; Cr = chromium; Mo = molybdenum; V = vanadium.

In Figs. 2, 3, and 4 are shown the magnetization curves for low, medium, and high magnetizing forces for these alloys, except the 80 permalloy and the 1.7 V-permendur, for which the curves are substantially the same as for 78.5 permalloy and permendur, respectively. Curves for "Armco" iron and ordinary commercial 4 per cent silicon steel also are shown in these figures. All these materials were annealed, and in the case of the 78.5 permalloy and the perminvars the annealing was followed by quenching and baking, respectively. It

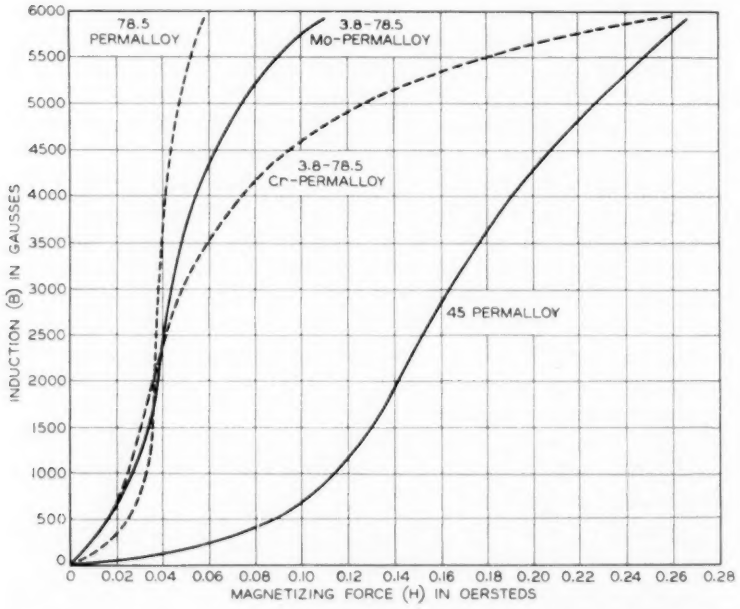


Fig. 2—Magnetization curves for several permalloys for flux densities less than 6,000 gausses.

may be seen from these curves that the permalloy group reaches almost saturation values long before the iron and silicon steel and the other alloys have reached the lower bend in the magnetization curve. With the exception of the 45-permalloy, which saturates at a fairly high value, the permalloys have low saturation induction and the permendur the highest. The permeability curves computed from these curves are plotted in Figs. 5 and 6. In Fig. 5, curves for the permalloy alloys are plotted at a smaller vertical scale than in Fig. 6 containing the curves for the other alloys.

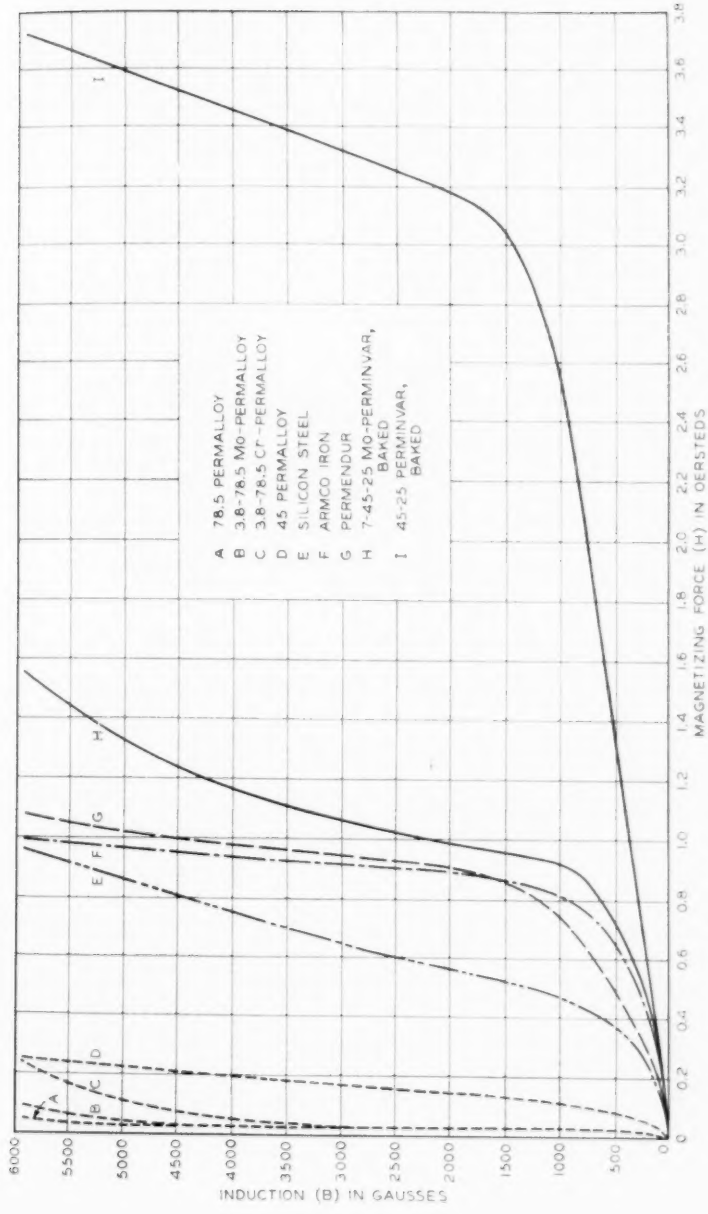


Fig. 3.—Magnetization curves for permalloys, perminvars, and permendur.

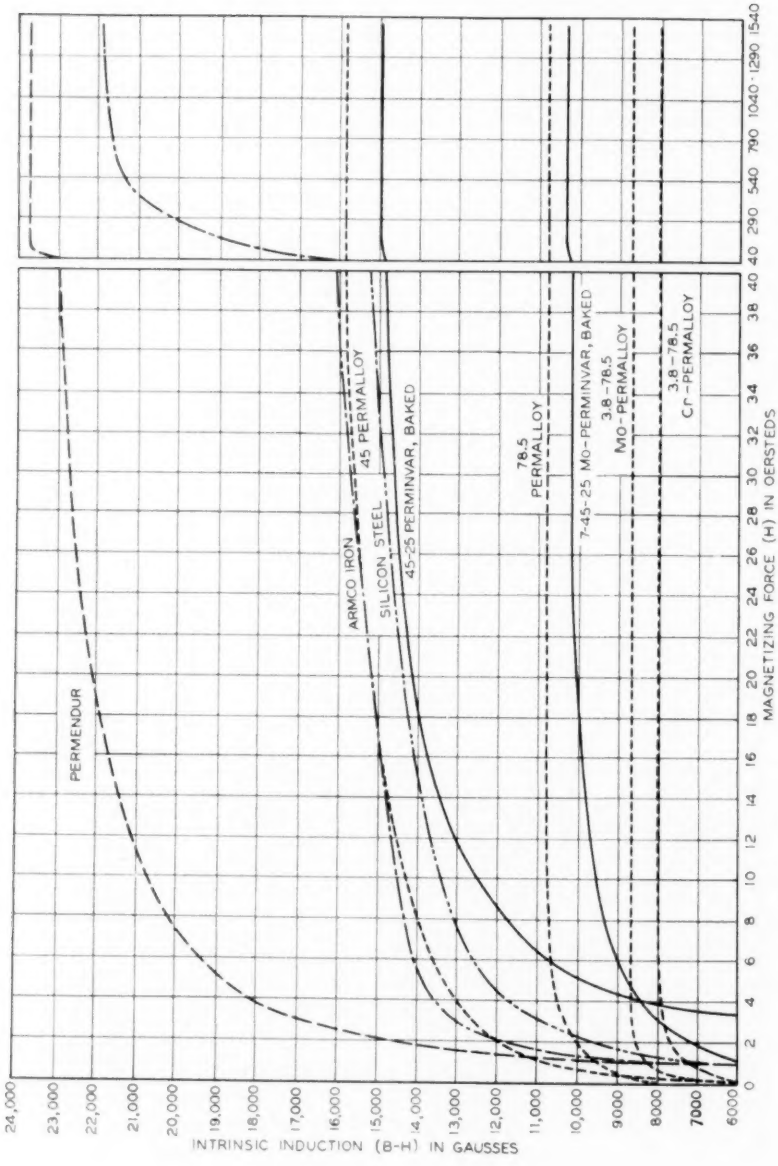


Fig. 4—Magnetization curves for permalloys, perminvars, and permendur.

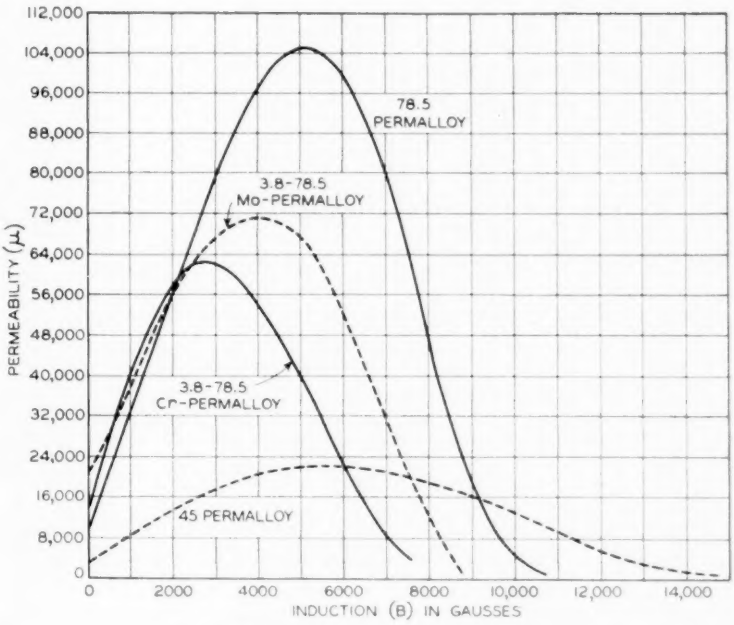


Fig. 5—Permeability curves for permalloys.

The permeability for alternating current of small constant amplitude as a function of superposed d.-c. magnetizing force is shown in Fig. 7 for some of the alloys. In most apparatus where both alternating and direct current are involved, this "butterfly curve" must be relatively flat over the expected range of d.-c. excitation. The important magnetic constants for these alloys are given in table II.

Permalloys

In Fig. 8 the initial and maximum permeabilities and the coercive force and resistivities are plotted for quenched alloys of the iron-nickel series. These curves show the remarkable variations in magnetic properties with composition in this series of alloys. The permalloy region includes alloys between 30 and 95 per cent nickel, as indicated in Fig. 1. Some of the alloys in this region, particularly from 50 to 85 per cent nickel, require rapid cooling to develop the magnetic properties indicated in the curves. If they are merely annealed, both the maximum and the initial permeabilities are much lower. The greatest effect in reducing the permeabilities by slow cooling appears to be for

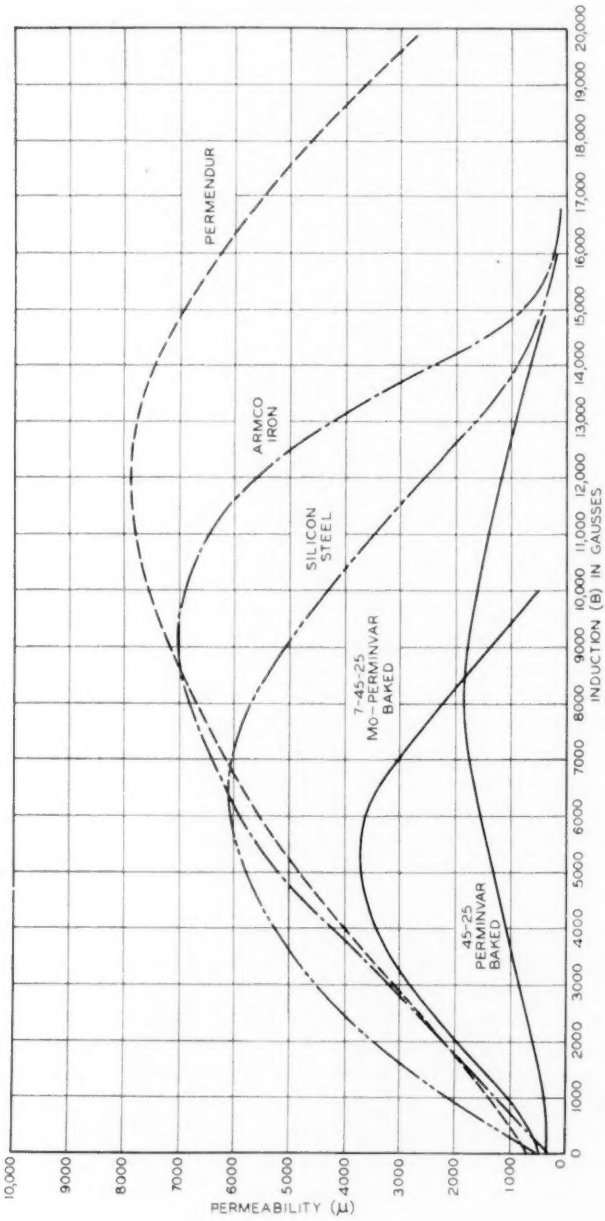


Fig. 6—Permeability curves for perminvars and permendur.

TABLE II

MAGNETIC CONSTANTS FOR ALLOYS DISCUSSED IN THIS PAPER

Material	μ_0	μ_m	$W_{H=\infty}$	B_r	H_c	$(B-H)_{H=\infty}$	ρ
"Armco" iron	250	7,000	5,000	13,000	1.0	22,000	11
4% silicon-steel	600	6,000	3,500	12,000	0.5	20,000	50
78.5 permalloy, quenched	10,000	105,000	200	6,000	0.05	10,700	16
45 permalloy	2,700	23,000	1,200	8,000	0.3	16,000	45
3.8–78.5 Cr-permalloy	12,000	62,000	200	4,500	0.05	8,000	65
3.8–78.5 Mo-permalloy	20,000	75,000	200	5,000	0.05	8,500	55
45–25 perminvar, baked	400	2,000	2,500	3,000	1.2	15,500	19
7–45–25 Mo-perminvar, baked	550	3,700	2,600	4,300	0.65	10,300	80
Permendur	700	7,900	6,000	14,000	1.0	24,000	6

Here μ_0 and μ_m are the initial and maximum permeabilities, respectively; $W_{H=\infty}$ is the hysteresis loss in ergs per cubic centimeter per cycle for saturation value of flux density; B_r is the residual induction in gausses; H_c is the coercive force in oersteds; $(B-H)_{H=\infty}$ is the saturation value of the intrinsic induction in gausses; ρ is the resistivity in microhms-centimeter.

the alloys containing between 70 and 80 per cent nickel; for example, 78.5 permalloy with a standard anneal has its initial permeability reduced to 1,200. If the alloy is baked for several hundred hours this permeability can be reduced still further to about 500. There is a very rapid decrease in the coercive force as the nickel increases above 27 per cent, and the lowest values are reached in the region between 70 and 80 per cent nickel. The resistivity increases rapidly just below the permalloy region, and reaches maximum at about 31 per cent nickel. It should be noted that the large changes in the coercive force and the resistivity are at the lower end of the permalloy region, while the highest permeabilities are developed in the alloys containing between 75 and 80 per cent nickel.

45-Permalloy

One of the alloys developed for commercial use is 45-permalloy. This attains a saturation flux density as high as any of the permalloys. At 40 oersteds the flux density is 16,000 gausses, substantially the same as for "Armco" iron, and considerably higher than for ordinary silicon steel (Fig. 4). The initial and maximum permeabilities (under standard practice of heat-treating) are 2,700 and 23,000, respectively (Fig. 5 and Table II). For cores requiring flux densities between 8,000 and 12,000 gausses this alloy is specially useful. The resistivity of the alloy is 45 microhms-centimeter, which is high enough to make it superior for use in cores in a.-c. circuits. The higher permeability at

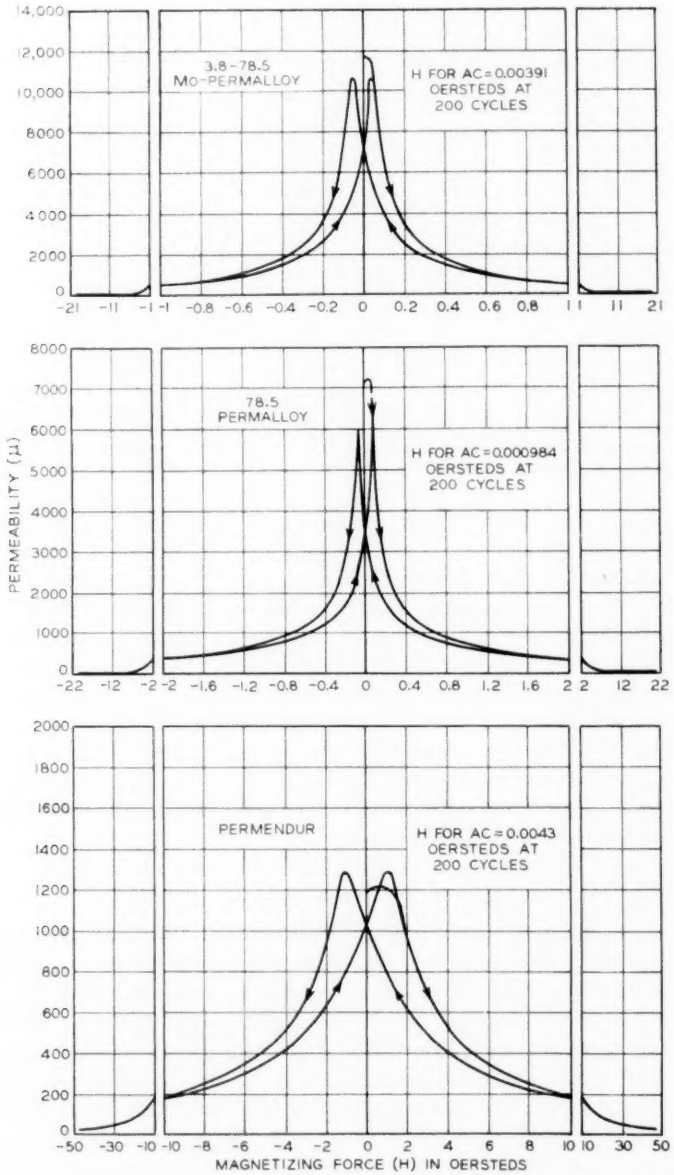


Fig. 7—Effect of superposed d.-c. fields on the a.-c. permeability of permalloys and permendurs. Annealed.—*Continued on page 125*

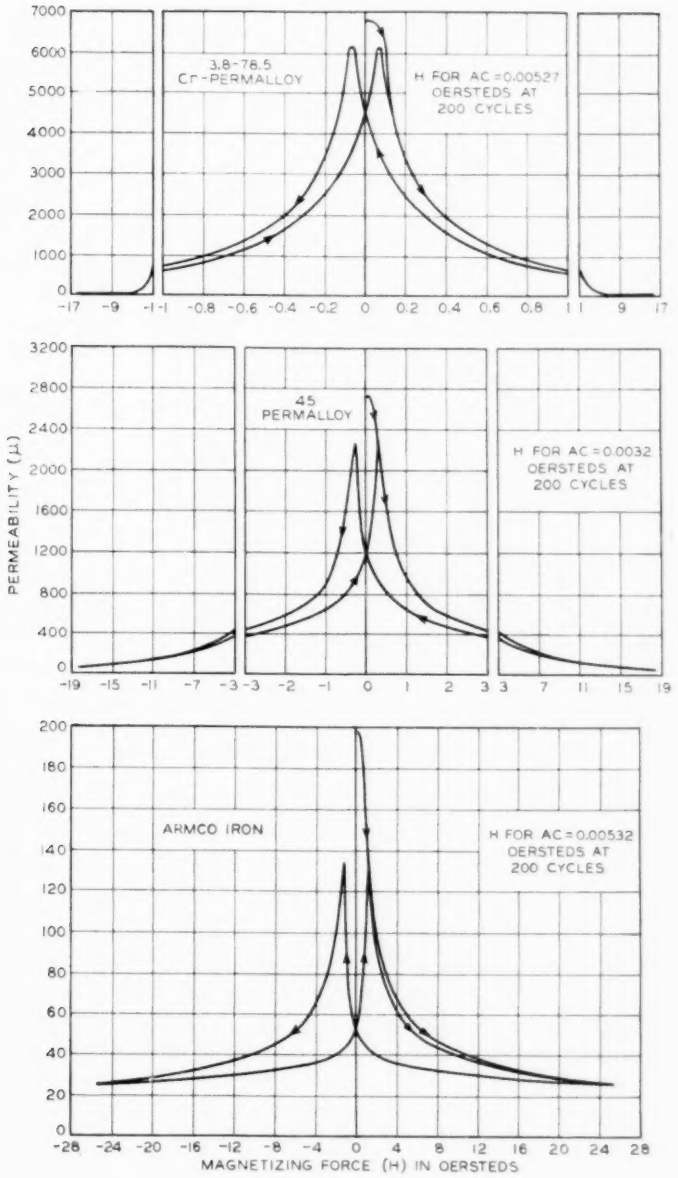


FIG. 7—*Continued from page 124*

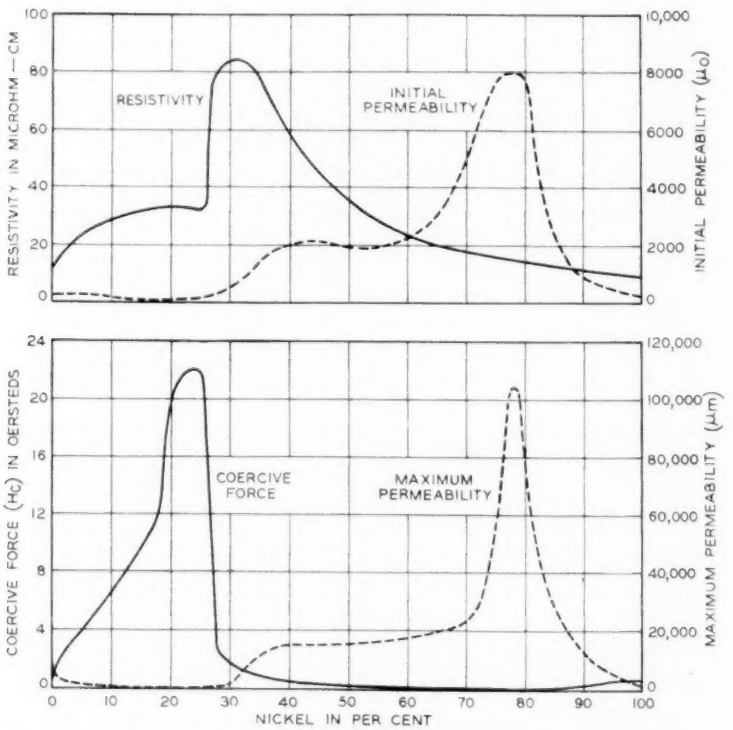


Fig. 8—Resistivity, initial and maximum permeabilities, and coercive force of iron-nickel alloys.

fairly high values of superposed d.-c. field, shown in Fig. 7, also favors its use for some purposes.

78.5-Permalloy

Another alloy long used in the telephone plant is 78.5 permalloy. Quenching develops a higher maximum permeability in this than in any of the other permalloys. Initial and maximum permeabilities of 10,000 and 105,000 readily are developed. The hysteresis loss and the coercive force of quenched 78.5 permalloy are minimum. The saturation flux density of this alloy is between 10,000 and 11,000 gausses, and it is reached with a very low magnetizing force. The rapid rise in the flux density of this alloy for small increments in the magnetizing force and the low saturation flux density are shown in Figs. 2, 3, and 4.

Initial and maximum permeabilities of 78.5 permalloy are improved by elimination of impurities and also by special care in the quenching process. As stated earlier, the rates of cooling required to develop the highest initial permeability differ from those for the highest maximum.

Chromium Permalloy and Molybdenum Permalloy

When other metals are added to permalloys their resistivities, in general, are increased. In research work at the Bell laboratories chromium and molybdenum mostly were used. It was found that with these elements a desirable combination of high resistivity and high initial permeability could be obtained. The variation in resistivity, keeping the nickel content constant at 78.5 per cent, is shown in Fig. 9. Chromium increases the resistivities somewhat more than molybdenum for a given addition, but the difference is not very large. The 3.8–78.5 Cr-permalloy has a resistivity of 65 microhms-centimeter, as compared with 55 for the 3.8–78.5 Mo-permalloy.

Figure 9 also illustrates the manner in which additions of these metals affect the initial permeability and the sensitivity of the permeability to rate of cooling. The solid line curves are for the annealed and the broken-line curves for the quenched specimens. For the quenched alloys the highest permeabilities are obtained when the added chromium and molybdenum are 2.4 per cent and 1.6 per cent, respectively. For this cooling rate the chromium permalloy seems to develop a slightly higher initial permeability. The difference, however, is small, and a greater spread between different samples has been observed. For the annealed alloys the largest value of initial permeability is obtained with molybdenum permalloy. For 3.8 Mo-permalloy an initial permeability of 20,000 is obtained. With the same heat-treatment the initial permeability of the corresponding chromium alloy is 12,000. It is surprising to note that small additions of these non-magnetic metals increase the initial permeability to values considerably higher than that for quenched 78.5 permalloy. Beyond 5 per cent this improvement ceases. All additions decrease the saturation induction values and the maximum permeabilities.

Several of these alloys have been developed for commercial use. Of these the most important are 2–80 Cr-permalloy, 3.8–78.5 Cr-permalloy, and 3.8–78.5 Mo-permalloy.

Perminvar

The distinctive magnetic properties of the perminvars are constancy of permeability at low flux densities, a low hysteresis loss in the same

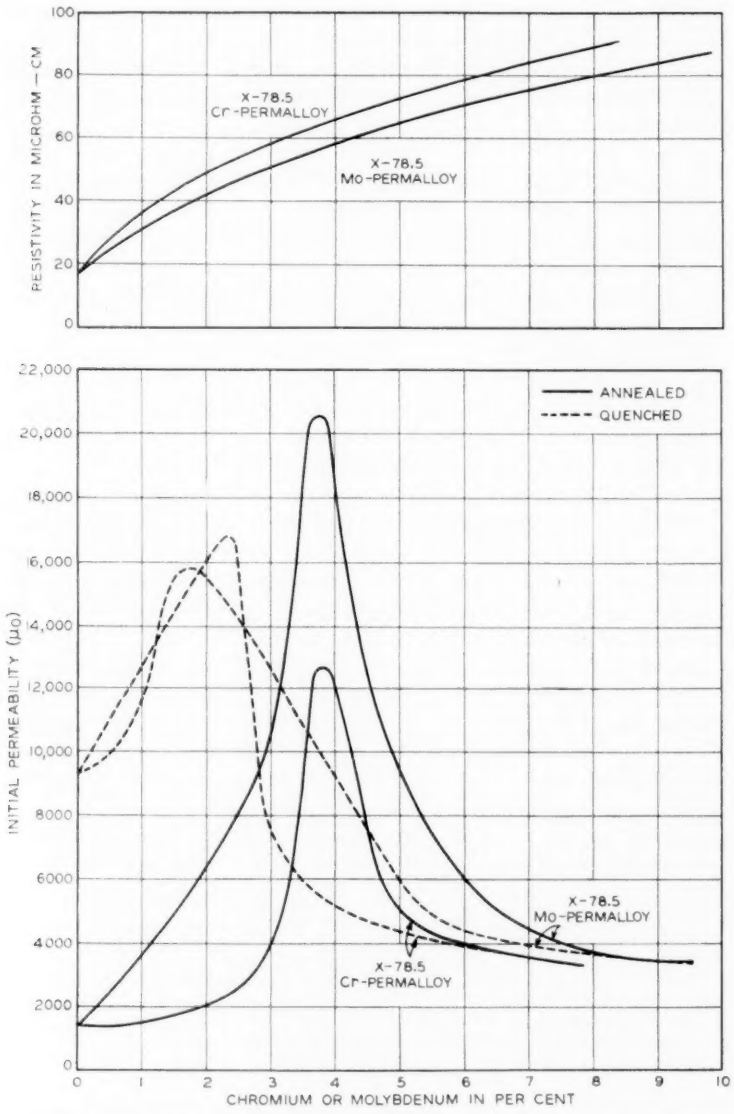


Fig. 9—Resistivity and initial permeability of 78.5 permalloy with chromium or molybdenum replacing part of the iron. In the curve designations, *X* indicates the percentage of chromium (Cr) or molybdenum (Mo) as read from the abscissa.

range, and, for medium flux densities, a characteristic constriction in the middle of the hysteresis loop. In some alloys this constriction is so extreme that the coercive force vanishes, making the two branches of the loop coincide when the magnetizing force is reduced to zero, in spite of the considerable hysteresis loss involved in the entire cycle. At high flux densities this constriction disappears and the loops have normal shapes.

The degree to which these properties can be developed depends on the composition and the heat-treatment. For the most typical alloys slow cooling in the annealing process produces this effect to a certain degree. Baking for 24 hours in the 400–500 degree (centigrade) temperature range brings most alloys into a stable condition in which no further baking materially will affect the magnetic properties.

As indicated in Fig. 1, some of the binary alloys tend toward the perminvar characteristics with long baking. Of the permalloys a considerable proportion of those that must be quenched to develop the desirable magnetic properties show perminvar characteristics when they are baked.

45–25 Perminvar

The perminvar characteristics have been developed most intensely in 45–25 perminvar. The magnetization curve in Fig. 3, and the permeability curve in Fig. 6, illustrate this fact. The constancy of permeability at low magnetizing forces and the necessity of "baking" to attain this condition are illustrated in one of the sections of Fig. 10, where the permeabilities are plotted for the quenched and baked conditions. The permeability of the quenched alloy begins to change at very low magnetizing forces, but that of the baked alloy, though lower, remains constant for magnetizing forces up to 3 oersteds.

Hysteresis loops for this alloy in the two conditions are shown in Fig. 10 for maximum flux densities of less than 1,000 and more than 5,000 gausses. For the baked alloy the hysteresis loops for maximum flux densities less than 1,000 gausses cannot be measured by ordinary ballistic methods, because the two sides of the loop coincide in a straight line. For loops with higher maximum flux densities the area begins to appear, but the two branches of the loop still meet at the origin. Although the coercive force is sensibly zero for the baked alloy until the maximum flux density exceeds 5,000 gausses, the hysteresis loss represented by the loop may become considerably greater than that for the quenched alloy.

7–45–25 Mo-Perminvar

The extremely low hysteresis loss and constancy of permeability at low flux densities makes 45–25 perminvar a suitable material for

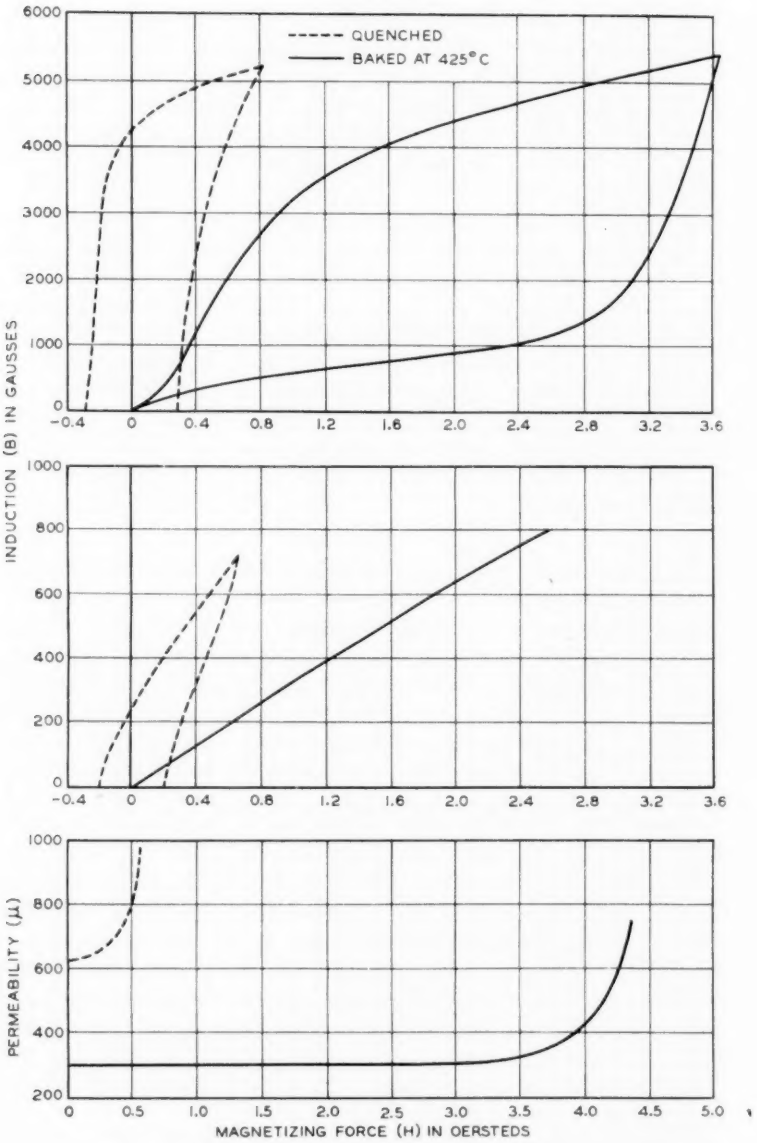


Fig. 10—Hysteresis loops and permeability curves for 45–25 perminvar.

applications where distortion and energy loss are fatal to good quality of transmission. The resistivity of this alloy is only 18 microhms-centimeter, but it can be increased without serious sacrifice of the low hysteresis characteristic by adding molybdenum. The alloy chosen for commercial use is 7–45–25 Mo-perminvar, having a resistivity of 80 microhms-centimeter.

The manner in which molybdenum affects the magnetic properties is illustrated in Figs. 3, 4, and 6. The permeability is not quite so independent of the magnetizing force as for the alloy without molybdenum, nor is the hysteresis loss quite so low. The initial permeability for the alloy baked the customary 24 hours is somewhat higher. When baked for a longer period the magnetic characteristics tend more toward those of 45–25 perminvar.

Permendur

An alloy in the iron-cobalt series used in communication apparatus is permendur. The typical composition is 50 per cent iron, 50 per cent cobalt. The outstanding magnetic property of this alloy is high permeability in the range of flux densities between 12,000 and 23,000 gausses (Figs. 4 and 6). The high permeability of this alloy "endures" to higher flux densities than does the permeability of any other magnetic material. Its initial permeability is about 700, though values as high as 1,300 have been observed for some samples. In addition to the high permeability at high flux densities permendur also has a relatively flat "butterfly" curve, as may be seen in Fig. 7. For a superposed d.-c. magnetizing force of 10 oersteds the a.-c. permeability of "Armco" iron is 40, as compared with 200 for permendur.

1.7 V-Permendur

Permendur is difficult to roll into sheets, because of its brittleness. To overcome this difficulty 1.7 per cent vanadium is added. With this addition it may be rolled into sheets as thin as a few thousandths of an inch. Th s amount of vanadium affects the magnetic properties only slightly, although larger amounts decrease the permeability at high flux densities.

Another improvement incident to the addition of vanadium is a fourfold increase of resistivity from its value of 6 microhms-centimeter for simple permendur. Permendur, it may be noted, has the lowest resistivity of the iron-cobalt series.

Laboratory Results

As stated hereinbefore, all the alloys that have been discussed have been made on a factory scale for some use in the telephone plant.

This paper, however, would be incomplete without mention of some of the remarkable magnetic properties obtained in laboratory samples. In table III some of these magnetic achievements are tabulated. The special treatments given these specimens are noted also in this table.

TABLE III

SOME REMARKABLE MAGNETIC PROPERTIES OBTAINED IN LABORATORY SPECIMENS *

Material	μ_0	μ_m	H for μ_m	H_c	Heat Treatment
"Armco" iron	20,000[1]	340,000[1]	0.021	0.03[1]	18 hr. at 1,480 deg. C., followed by 18 hr. at 880 deg. C., both in hydrogen
45 permalloy	11,000	227,000	0.025	0.0145	Melted in vacuum; electrolytic iron and electrolytic nickel; 18 hr. at 1,300 deg. C. in hydrogen
65 permalloy	2,500	610,000[2]	0.0148	0.012[2]	18 hr. at 1,400 deg. C. in hydrogen; heated to 650 deg. C. 1 hr., cooled in magnetic field of 16 oersteds in hydrogen
78.5 permalloy	13,000[3]	405,000[3]	0.0101	0.0153	$2\frac{3}{4}$ x 2 x 0.109 in. tape; annealed; wrapped in 2 layers of 3 mil. tape and quenched from 600 deg. C. in tap water
3.8–78.5 Mo-permalloy	34,000[1]	140,000[1]	0.025[1]		1,400 deg. C. in hydrogen
Permendur	1,000	37,000	0.22	0.20	940 deg. C. in hydrogen for 18 hr., slowly cooled to room temperature
Permendur	1,300	29,000	0.27		940 deg. C. in hydrogen, 6 hr.; slowly cooled to room temperature
45–25 perminvar		189,000	0.052	0.059	Heated to 1,000 deg. C. in hydrogen, reheated to 700 deg. C. and cooled in hydrogen in a magnetic field of 14 oersteds

For explanation of symbols in headings see footnote to table II.
* Except as noted, the values in this table have not previously been published.

ENGINEERING APPLICATIONS

One of the first uses of the permalloys was for continuous loading of a telegraph cable between New York and the Azores laid in 1924. For this project 78.5 permalloy was used in the form of a 0.125 by 0.006 inch tape wrapped helically on a stranded copper conductor. The average initial permeability of this alloy in the laid cable was 2,300,

considerably less than can be obtained under the best conditions of heat-treatment and absence of strains. With this loading, the speed of transmitting messages was increased fivefold.[4] By the time a second cable project was undertaken the chromium permalloys had been developed, and 2–80 Cr-permalloy was selected. This alloy has a resistivity of 45 microhms-centimeter, and the initial permeability of the loading on the laid cable was in the neighborhood of 3,700. The increase in permeability and in resistivity increased materially the message carrying capacity.[5]

The largest use of permalloys in the telephone plant has been in cores of loading coils,[6] where the alloy is used in the form of compressed insulated dust. Iron dust cores had been standard for these coils.[7] The lower magnetic losses of permalloy dust, however, permitted utilizing higher core permeabilities. This has resulted in a very material decrease in the size of loading coils. For a high grade loading coil core made from iron dust the effective core permeability at low flux densities had to be limited to 33. The first permalloy used for loading coil cores was 80 permalloy. The insulated and compressed core was designed for an effective permeability of 75—more than double that for the iron dust. Development work on an improved compressed magnetic dust core in which molybdenum is used, is now approaching completion. It is expected that the new material will have a substantially higher permeability than that of the 80-permalloy dust cores, and that it will have intrinsically superior eddy current and hysteresis loss characteristics. By virtue of these properties, it will be practicable to make a further substantial reduction in the size of loading coils without sacrifice in service standards. The decrease in the size of the cores with improvement in the core material is illustrated in Fig. 11.

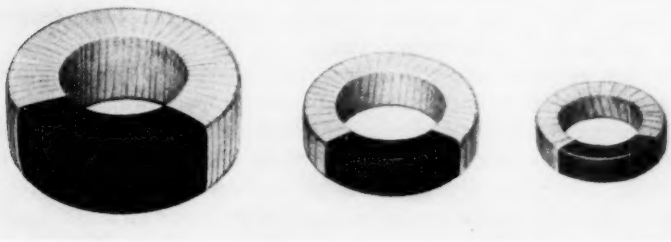


Fig. 11—Equivalent cores for loading coils. Iron dust core (left); permalloy dust core (center); molybdenum permalloy dust core (right).

[4] For all numbered references see list at end of paper.

In d.-c. apparatus where high permeability and low coercivity are of importance, and where high resistivity does not add to the usefulness of the core material, 78.5 permalloy is suitable. It is used in certain relay structures, usually of marginal type, in which the difference between operating and releasing currents is small.

For audio transformers, for retardation coils, and for other apparatus in which high permeability and high specific resistance must be combined, both 3.8–78.5 Cr-permalloy and 3.8–80 Mo-permalloy have been used. The former has slightly higher resistivity, but the latter has higher initial permeability and is more ductile.

While the initial and maximum permeabilities of 45-permalloy are not as high as those of 78.5-permalloy, the higher flux densities attained by the former and its higher resistivity favor its use for certain types of relays and transformers where high flux densities are required. It is used also in some instances for cores of coils that require high a.-c. permeability when d.-c. magnetizing forces are superposed.

The magnetic characteristics of the perminvars make them especially suitable for use in circuit elements in which distortion and energy loss must be a minimum; but their relatively high cost, and the advisability of avoiding high magnetization throughout the life of the apparatus, have prevented their extensive use in telephone plant. One use for which perminvar is especially suitable is the loading of long submarine telephone cables. Here a high resistivity is very desirable, which has been shown to be obtainable in the 7–45–25 Mo-perminvar. The increase in resistivity resulting from the addition of molybdenum more than offsets the accompanying increase in hysteresis loss, and results in a continuous loading material satisfactory for certain types of loaded cables.

Permendur was developed for use in apparatus where very high flux densities are desired. For a moderate magnetizing force flux densities of 18,000 and 23,000 gausses readily are obtained. It is used for cores and pole pieces in loud speakers, certain telephone receivers, light valves, and similar apparatus.

It may be seen from this survey that there is a great variety of magnetic materials with widely different properties from which an engineer may choose in designing magnetic elements in which magnetic flux changes are essential. Already these alloys have an important place in telephone plant. However, iron and silicon steel still are used extensively, and will continue to hold their own on a cost basis for some purposes. There is no doubt, however, that alloys of iron, nickel, and cobalt will continue to supplant iron and silicon steel in many places where circuits and apparatus are redesigned to take full advantage of their magnetic properties.

References

1. Unpublished results, P. P. Cioffi.
2. J. F. Dillinger and R. M. Bozorth. Heat Treatment of Magnetic Materials in a Magnetic Field. *Physics*, Vol. 6, September, 1935, p. 279–91.
3. G. W. Elmen. Magnetic Alloys of Iron, Nickel, and Cobalt. *Franklin Inst. Journal*, Vol. 207, 1929, p. 583–617.
4. O. E. Buckley. The Loaded Submarine Telegraph Cable. *A. I. E. E. Journal*, Vol. 44, August 1925, p. 821–9.
5. O. E. Buckley. High Speed Ocean Cable Telegraphy. *Bell System Technical Journal*, Vol. 7, 1928, p. 225–67.
6. W. J. Schackelton and I. G. Barber. Compressed Powdered Permalloy, Manufacture and Magnetic Properties. *A. I. E. E. Transactions*, Vol. 47, April, 1928, p. 429–36.
7. B. Speed and G. W. Elmen. Magnetic Properties of Compressed Powdered Iron. *A. I. E. E. Transactions*, Vol. 40,1921, p. 1321–59.

Improvements in Communication Transformers *

By A. G. GANZ and A. G. LAIRD

The rapidly advancing art of electrical communication and the increasingly wide variety of its applications have required marked improvements in the transformers used in communication circuits. These improvements, achieved partly through advances in design and partly through improvements in the constituent materials, are discussed in this paper.

THE rapid development of the art of electrical communication in the last decade has necessitated marked improvements in the transformers used in it. New applications for these transformers and the extension of old ones have imposed new and far severer performance requirements. The primary applications of communication transformers are in the telephone plant, in the various voice and carrier transmission circuits, and in a multitude of incidental services. They have also wide uses in radio broadcasting transmitters and receivers, in the amplifiers of sound motion picture equipment, in the radio equipment for aircraft, and in a variety of other circuits.

Although communication and power transformers have a common origin, the communication transformer now has evolved as a precision device which has only a general resemblance to the usual power transformer. Some voice-frequency transformers, such as those used in aircraft, weigh but 2 or 3 ounces, yet transmit speech substantially undistorted. Some used in program circuits transmit with negligibly small phase or amplitude distortion all frequencies from 20 to 16,000 cycles per second. Transformers also have been developed for transmitting narrow bands of frequencies and having associated with the normal transformer performance valuable frequency discriminating properties. A discussion of improvements in these narrow band transformers is outside the scope of this paper, which will be confined to those transmitting wide-frequency bands, that is, those for which the ratio of upper to lower limiting frequencies is at least 10 to 1.

The design of the modern communication transformer is based upon extensions of the familiar theory of transformers covered in numerous texts. However, this type of transformer is a more complex device, with its multiplicity of requirements and its transmission over a wide-frequency range. Its proper representation accordingly requires a

* Published in the December 1935 issue of *Electrical Engineering*, and scheduled for presentation at the Winter Convention of the A. I. E. E., New York City, January 28–31, 1936.

more elaborate equivalent network than the customary Π or *T* network. With the use of such a network, the performance of the transformer may be correlated mathematically with its constants in accordance with network theory.

Improvements in Low-Frequency Transmission

The great improvements in communication transformers, particularly at audio frequencies, are largely attributable to the invention and application of the permalloys as magnetic core materials. For convenience, the term "permalloy" has been applied to a group of nickel-iron alloys containing between 30 and 95 per cent nickel which have been developed by Bell Telephone Laboratories.[1, 2, 3] In addition to other desirable magnetic properties, some of these permalloys [4] when properly heat treated yield exceptionally high initial permeabilities. As a result of the use of these special alloys, telephone transformers may be designed to have less loss and distortion over wider frequency ranges than has been possible in transformers designed without the benefit of use of such materials.

Figure 1 illustrates the excellence of performance resulting from the use of a special permalloy consisting of approximately 4 per cent

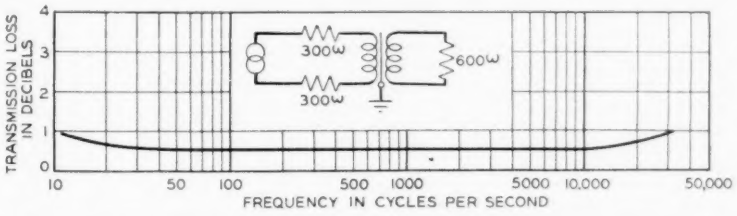


Fig. 1—Transmission-frequency characteristic of a transformer utilizing a permalloy core and designed to connect a telephone transmission line to a program repeater. Transmission loss shown is the loss relative to an ideal transformer of the same ratio.

chromium, 78 per cent nickel, and the remainder iron, in a transformer designed to connect a telephone transmission line to a program repeater. Figure 2 shows the voltage amplification characteristic of an interstage transformer for a high quality amplifier. As is well known,[5] superimposed direct currents generally decrease the effective a.-c. permeability of ferromagnetic materials. Therefore, to retain the full benefit of the permalloy core, an auxiliary circuit was used with this transformer for supplying the plate current to the preceding tube. Another illustration is a transformer designed for use as an input transformer (that is, one designed to operate into the grid circuit of a

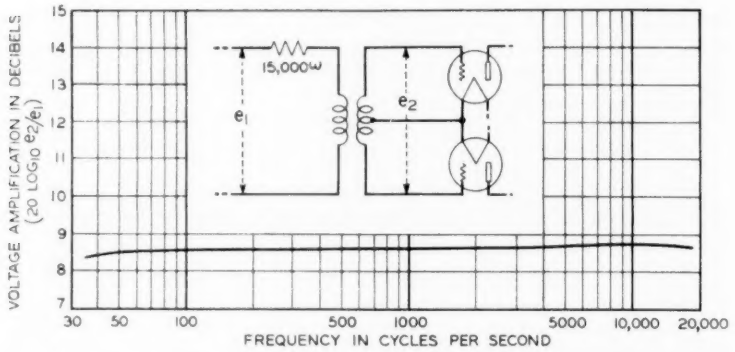


Fig. 2—Voltage amplification-frequency characteristic of an interstage transformer for a high quality program amplifier.

vacuum tube) for portable recording equipment where light weight is an important consideration. The voltage amplification-frequency characteristic of this transformer is shown in Fig. 3. There is shown also in this figure, for purposes of comparison, the very much poorer characteristic realized when an alloy of about 50 per cent nickel as commonly used is substituted for the special permalloy. In addition,

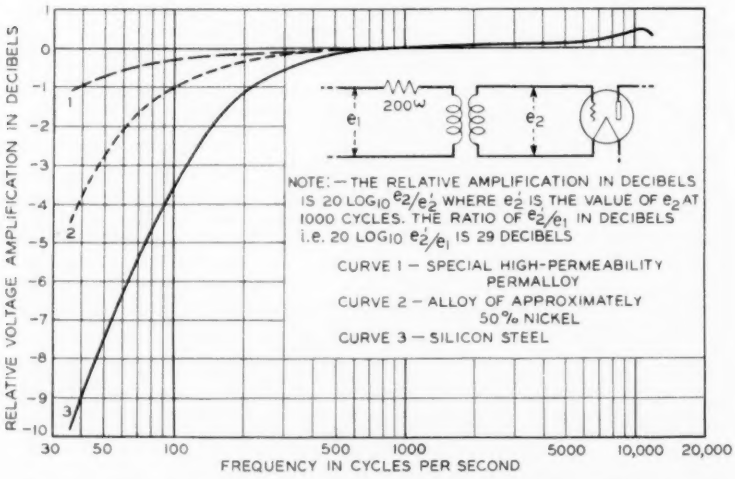


Fig. 3—Curves showing the relative effect of different core materials on the voltage amplification-frequency characteristic of a small transformer designed for portable recording apparatus. The composition of the special permalloy is approximately 4 per cent chromium, 78 per cent nickel, and the remainder iron.

there is shown the effect of using silicon steel as core material, which was the practice in older transformers.

An important limitation of high permeability alloys is their sensitivity to mechanical strain which may seriously impair their magnetic characteristics. Considerable care must be exercised to avoid strain during assembly operations after laminations are annealed. Telephone transformers are designed specially to provide a firm assembly without mechanical strain, thereby retaining the high permeabilities available.

Increase in Voltage Amplification

As may be seen from the foregoing curves, the voltage amplification of input transformers at the low end of the frequency band is directly dependent on the permeability of the magnetic core material. At the highest frequencies the voltage amplification of the above input transformers is controlled by leakage and capacitances, the latter including grid circuit capacitances as well as the transformer distributed capacitances. Over a wide range in the central part of the frequency band these effects are negligible and the transformer performs much as an ideal transformer of the same ratio. By proper proportioning of the leakage and capacitance effects, the shape of the characteristic may be controlled to a certain measure at will. For example, a rising voltage amplification-frequency characteristic can be obtained if desired to correct for a falling characteristic of other parts of the amplifier.

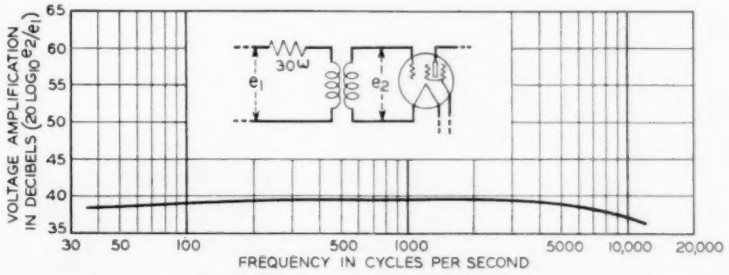


Fig. 4—Voltage amplification-frequency characteristic of an input transformer having an impedance ratio of 10,000 to 1.

In certain types of circuits the voltage amplification of input transformers is at a high premium, such as in the amplification of low-energy signals when a.-c. power is used for the tubes. Under these conditions the tubes tend to introduce appreciable noise. A high-voltage amplification in the input transformer serves to raise the signal voltage at the grid terminals so as to override the tube noises. Figure 4 shows

that a high amplification, in this instance 10,000 to 1 in impedance level from a 30-ohm source, may be realized without undue restriction in either the low- or the high-frequency transmission.

Since transformers of this type are located at points of very low energy levels, special pains must be taken to avoid interference from stray magnetic and electrostatic fields. To prevent hum from nearby power apparatus, transformers are enclosed in cases or shields of high permeability material. The interference voltages induced in these transformers are some 30 or 40 decibels less than in older unshielded types of transformers. For higher frequency interference, effective shielding is obtained by cases made of high-conductivity material such as copper or aluminum.

Reduction in Size and Weight

The demand for lightweight equipment for aircraft communication and for portable apparatus for testing and recording has resulted in the development of communication transformers of unusually small size and weight. The smaller sizes weigh only $3\frac{1}{2}$ ounces and occupy a space of but 3 cubic inches. One of these used in aircraft receiving sets is illustrated in Fig. 5 contrasted with an earlier transformer also for lightweight service. The transmission loss-frequency characteristics are shown in Fig. 6. The corresponding characteristic of an input transformer for similar service is shown in Fig. 7.

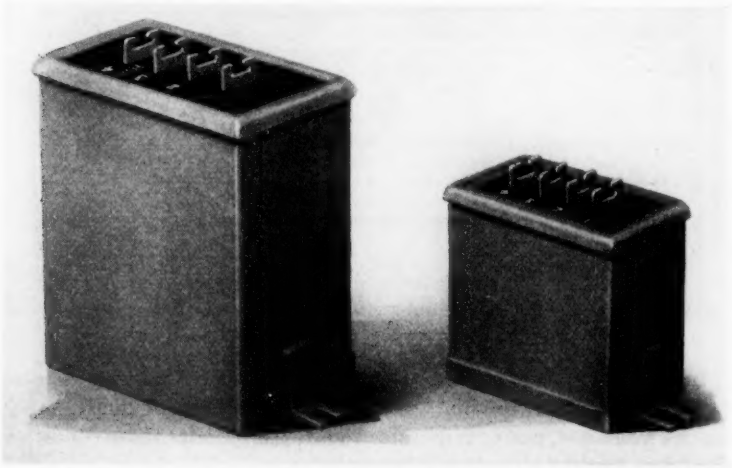


Fig. 5—Output transformer *A* (right) utilizing a permalloy core transmits frequencies from 40 to 3,000 cycles with greater over-all efficiency than the larger output transformer *B* (left) utilizing a core of silicon steel (see Fig. 6).

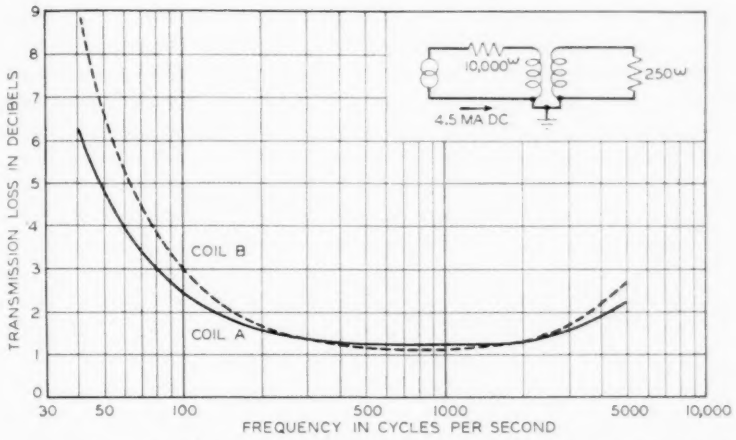


Fig. 6—Curves showing the transmission loss-frequency characteristics of (*A*) the small output transformer and (*B*) the larger output transformer shown in Fig. 5.

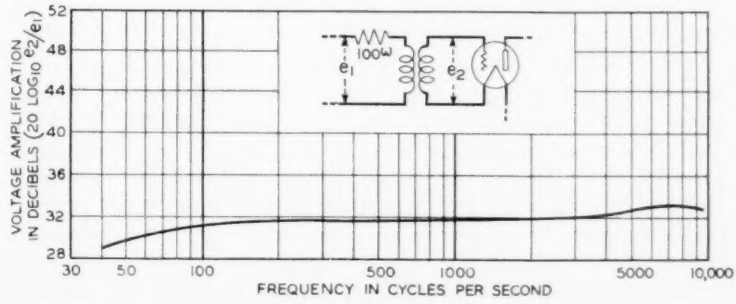


Fig. 7—Voltage amplification-frequency characteristic of an input transformer similar in size to the smaller output transformer shown in Fig. 5.

Extension of Range to Higher Frequencies.

The effective permeabilities of alloys of high initial permeabilities drop rather rapidly with frequency, a property which lessens the value of such alloys in transformers for carrier and higher frequencies. This effect, which is attributable primarily to eddy currents, can be greatly decreased, of course, by the use of thinner laminations. The use of these alloys, however, is limited by the rapidly increasing cost of reducing the lamination thickness and the less efficient use of the volume available for the core. This dropping off of effective permeability with frequency is not so important in audio-frequency trans-

formers, since there the core characteristics limit the transmission only at low frequencies where the effective permeability is high.

At higher frequencies the voltage amplification is severely limited also by the grid circuit and transformer capacitances. It has been found advantageous to add correcting elements, such as inductances and capacitances, to increase the gain ordinarily available. These added elements and the equivalent elements in the transformer are designed together as configurations similar to low-pass filter sections terminated midshunt in the grid circuit capacitance. Figure 8 illus-

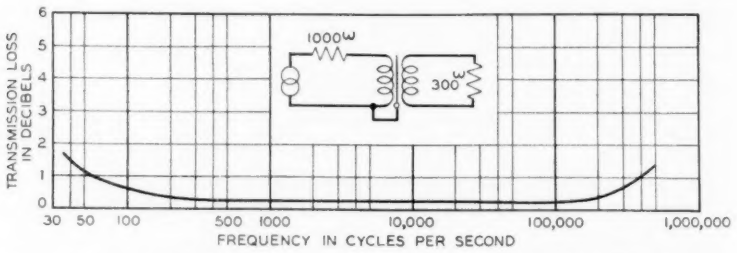


Fig. 8—Transmission loss-frequency characteristic of a transformer designed to transmit high frequencies in addition to audio frequencies.

trates the performance of a transformer designed for certain high frequency transmission experiments. The performance of this transformer in the frequency range of 30–500,000 cycles per second is similar to that of earlier types in the range of 30–60,000 cycles per second.

In most telephone applications the design of input transformers is complicated greatly by circuit functional requirements in addition to those of direct transmission. For example, as discussed in a succeeding section of this paper, phase distortion requirements demand self-impedances far higher than those required by transmission loss considerations, thermal [6] noise requirements demand lower dissipations, and impedance requirements limit the voltage amplification obtainable. For feed-back type amplifiers [7] the input transformers in addition to the forward amplification must meet transmission and phase requirements in the regenerative path. These special requirements apply not only for the transmitted frequency band but also at frequencies remote from this band in order to insure stability of the amplifier.

Reduction in Phase Distortion

Since transformers are reactive devices, they introduce phase shift in the circuits in which they are used. If the phase shift introduced be a linear function of the frequency it will not produce any distortion

in the shape of the transmitted wave. However, departures from linearity change the wave shape, and this form of distortion is referred to as phase or delay distortion. The delay at any frequency is a measure of this departure from linearity, and is dependent upon the frequency derivative of the phase shift at that frequency. Differences in delay of the various frequency components of the signal wave which transformers tend to produce result in distortion that may be especially serious in circuits intended for program transmission.

For wide-band transformers the delay caused by the shunting effect of the mutual impedance usually predominates. In fact for audio transformers the delay at higher frequencies is relatively so small that the delay distortion is practically equal to the mutual impedance delay at the lowest transmitted frequency. Delay distortion is also of importance in transformers to be used in television and telephotography. In these circuits phase distortion causes a space shift in the image of certain frequency components with respect to others with consequent blurring of the image.

The delay characteristic of a transformer used in program circuits to connect a telephone transmission line to the grid circuit of a repeater amplifier is shown in Fig. 9. This characteristic is compared in the same illustration with the delay characteristic of a repeater transformer developed some years ago for use in what then was re-

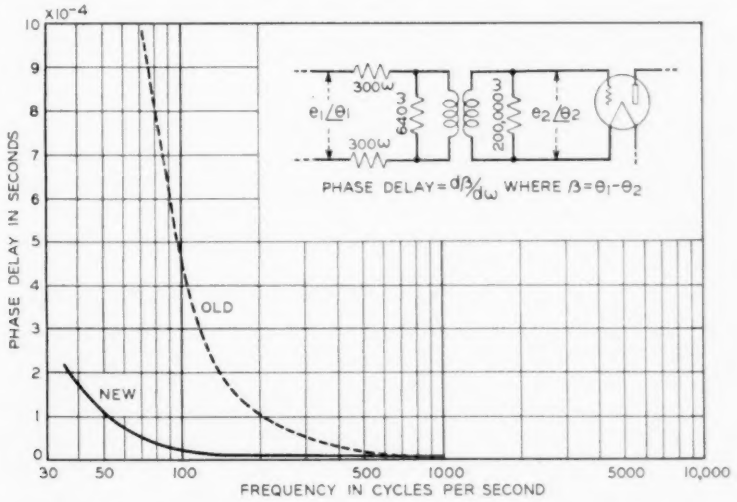


Fig. 9—Phase delay-frequency characteristic of an input transformer used in recent program repeaters, compared with that of an input transformer used in an older repeater.

garded as a high quality circuit. The delay characteristic of a high frequency transformer is shown in Fig. 10.

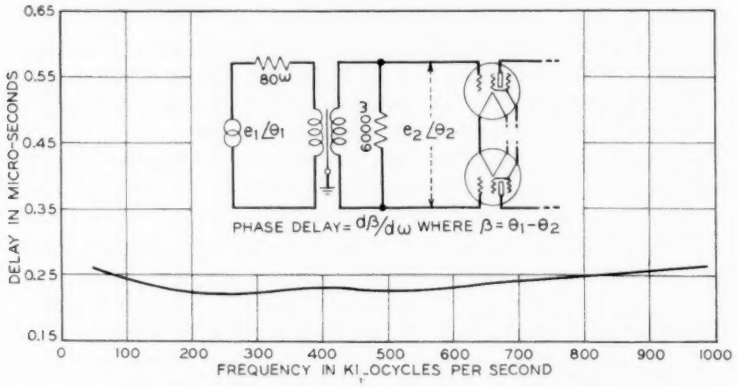


Fig. 10—Phase delay-frequency characteristic of an input transformer designed to transmit radio frequencies.

Reduction in Audio Frequency Modulation

The present exacting requirements for transformer performance have made it necessary to lessen greatly certain second-order distortion effects inherent in transformers having magnetic cores. Nearly all core materials tend to generate extraneous frequencies because of magnetic non-linearity—a property referred to as magnetic modulation. In audio-frequency circuits intended for high quality service, magnetic modulation may cause serious distortion in that harmonics of the lower frequencies appear higher in the audible range. The energy present at the lower frequencies is usually so much greater than that over the rest of the band that the modulation products may approach the order of magnitude of the signal components at higher frequencies.

This form of distortion in no way is revealed by the ordinary transmission loss characteristic; in fact, a transformer having a very flat loss characteristic over a wide-frequency range may nevertheless be definitely objectionable from a modulation standpoint. In present audio-frequency transformers, the total modulation products are some 40 to 80 decibels down from the energy of frequencies around 35 cycles per second producing them. This represents an improvement of about 30 decibels over older types.

Another second-order effect resembling modulation is microphonic noise caused by magnetostriction phenomena, that is, changes in magnetization accompanying the physical deformation of the magnetic

material. For instance, slight jarring of input transformers used in very high-gain amplifiers (100 decibels or more) may induce in this manner disturbing noise voltages. Freedom from magnetostriction is a unique characteristic of permalloys containing approximately 80 per cent nickel, and their use accounts for the superiority of telephone transformers from this standpoint.

Reduction in Carrier Frequency Modulation

Magnetic modulation is even more serious at carrier frequencies where a transformer may transmit many channels, frequently at widely different levels. The modulation from the higher level channels may produce very objectionable interference in other channels. Carrier transformers now have been improved to such an extent that highly sensitive testing circuits are required to detect and measure the modulation products contributed by them. Representative values of these modulation products expressed as current ratios to the fundamentals are of the order of one-millionth of the fundamental frequencies, compared to one-thousandth in older types.

It is of interest to point out that in such transformers the presence of magnetic material, other than the special permalloys, in the vicinity of the transformer must be avoided with great care. For example, a small steel screw near the field of the transformer will seriously impair its performance from a modulation standpoint. Common practice is to use brass parts for the assembly and to confine the field of the transformer by completely enclosing it in a copper or aluminum case. The transformer then may be mounted by any convenient means without affecting its performance.

Improvements in Shielding and Balance

In the very nature of the service that it renders, the telephone plant involves many independent communication circuits in fairly close proximity. The minimizing of interference between these independent communication circuits has constituted a major problem in telephone engineering. Where such interference occurs between like circuits, that is, between two voice circuits or between two carrier circuits of overlapping frequency bands, the interference commonly is referred to as crosstalk, as distinguished from the interference from other types of circuits such as power and telegraph circuits.

In order to avoid crosstalk and other interference, balanced * cir-

* The new coaxial cable circuits under development are an interesting exception. In this type of cable a grounded outer conductor completely encloses the central conductor, and shielding rather than balance is relied upon to protect the circuit from interference. The shielding depends upon the size, thickness, permeability, and conductivity of the outer conductor and the frequency of the disturbance. If the frequency band used in the transmission over coaxial circuits is chosen properly, the circuit may be made substantially immune to effects from outside fields.

cuits are used almost exclusively in the telephone plant. For simplicity, some of the terminal apparatus connected with these circuits is constructed unbalanced, so that it is necessary to interpose transformers between the lines and the office equipment, these transformers providing a barrier to the propagation of the relatively large longitudinal currents from the line circuits. (Longitudinal currents, in contrast with the usual circulating currents, are currents that flow equally and in the same direction in both sides of the line.) In order to insure that the voltage impressed on the office equipment is attributable to the voltage between the wires of the line circuit and not to that between the wires and ground, it is necessary that the transformers be balanced very carefully; and for certain types of circuits, shields must be interposed so that the direct capacitance between the line winding and office winding is reduced to a very small value.

With a greater emphasis on carrier frequency transmission, a higher degree of balance is required between certain transformer windings, and highly effective shielding is frequently necessary. It is necessary also that the line windings be balanced very closely with respect to capacitances to the shield and case. The unbalance effects in carrier transformers now have been reduced to values in the order of 1 or 2 microamperes in circulating current per volt between the line windings and ground at 30,000 cycles per second, which compares with values of 50 microamperes or more for older transformers. At the same time the electrostatic shielding between the windings has been improved to such an extent that the direct capacitance between windings has been reduced to 1 or 2 micromicrofarads instead of 30 or 40 as before. The shields are arranged to intercept the dielectric flux lines tending to connect the primary and secondary windings, so that the direct capacitance between the two windings is attributable only to stray flux which bypasses the shield. One of the windings usually is enclosed completely in lead or copper foil with overlapping edges insulated to prevent a short-circuited turn. Still further improvements are obtained by covering the leads with grounded metal braiding, and in special cases by enclosing the terminals of the shielded winding in a separate shielded compartment. In certain transformers designed for high precision testing equipment, the direct capacitance between windings has been reduced to values less than 0.001 micromicrofarad.

In connection with phantom circuits, severer crosstalk requirements have necessitated more precise balances in the associated voice frequency transformers. In these transformers the turns are so arranged that the various distributed capacitances, flux linkages, and d.-c. resistances are disposed symmetrically with respect to ground. It has

been found in practice that this symmetry is realized most readily by close coupling between the various parts of the windings. By improvements in design, the crosstalk between phantom and side circuits has been reduced to values in the order of 20 millionths in current ratio, compared to values 5 times as large, formerly tolerated.

Reduction in Impedance Distortion

As a further consequence of the extension of carrier systems, it has become necessary to match the impedance presented by transformers, when terminated in the succeeding circuits, to particular values over the frequency range. For example, the transposition schemes used on open wire lines are such as to minimize crosstalk primarily for carrier signals propagated in one direction in any line. If the transformer terminating such a line does not present an impedance under load equal to the characteristic impedance of the line, a portion of the wave is propagated in the reverse direction, that is reflected, causing crosstalk into adjacent circuits. This reflection effect increases with the vector difference in the impedances of the transformer and the line, the latter impedance approaching a pure resistance as the frequency increases.

The impedance of transformers has become increasingly important where such transformers terminate filters that require a nearly constant resistance termination to maintain proper attenuation characteristics. Another example is in transformers terminating screen grid tubes where the plate impedances are relatively very high. Here the energy abstracted from the plate circuit and transmitted by the transformer is directly dependent upon the resistance component of the impedance of the transformer when terminated in its load.

Better impedance characteristics of transformers for these various applications have been obtained by increasing the mutual impedance and decreasing the leakage and capacitance effects. This procedure is made difficult by the necessity for meeting at the same time other and newer requirements, as, for example, modulation limits. Correcting elements consisting of capacitances and inductances usually are added and are proportioned with the transformer elements in accordance with network theory. Typical impedance characteristics of such transformers are shown in Figs. 11 and 12.

Input transformers operating into the grid circuits of vacuum tubes inherently have impedances that depart widely from the nearly pure resistances usually desired, because of the reactive termination provided by the grid circuit. This makes it necessary to add resistances to serve in place of the usual load resistance. The required dissipation

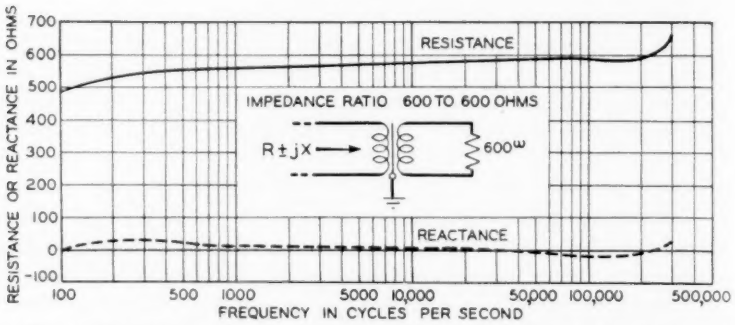


Fig. 11—Impedance-frequency characteristic of a transformer designed to operate between 600-ohm terminations and to transmit high frequencies in addition to audio frequencies.

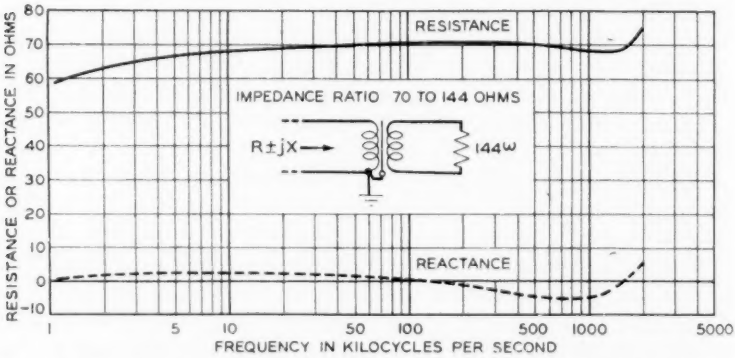


Fig. 12—Impedance-frequency characteristic of a transformer designed to operate between resistance terminations of 70 ohms and 144 ohms.

may be provided in many ways in the input transformer, which consideration allows much wider latitude in the design than in the ordinary transformer; there the major part of the dissipation for low transmission loss necessarily must occur in the load. A typical impedance characteristic of such an input transformer is shown in Fig. 13.

Testing Precautions

As a necessary concomitant to improvements in transformers, more precise testing circuits have been developed for accurately determining transformer performance. In estimating the performance of the transformer from its characteristic, care must be taken to make sure that the service conditions were reproduced carefully in the measuring circuit. In particular, certain precautions must be observed in the

measurement in order to avoid obtaining a misleading characteristic. For example, the permeability of magnetic core materials tends to rise rapidly from its initial value with increasing voltages. If in the measurement, the low-frequency voltages used are materially higher than they are under service conditions, the low-frequency response will appear to be much better than the true response.

As another example, the transmission of input transformers at the high-frequency end may be critical with the termination of the high-

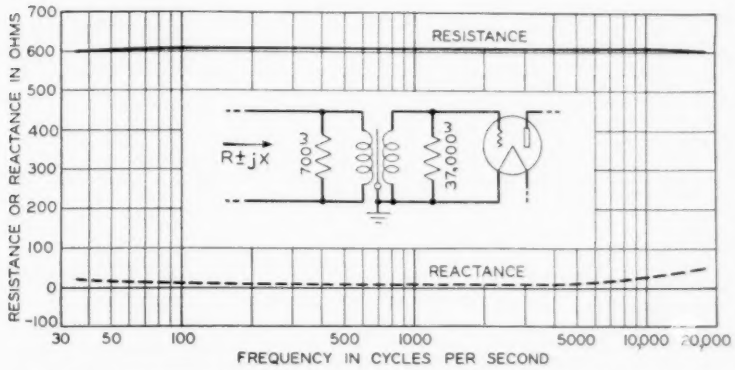


Fig. 13—Impedance-frequency characteristic of an input transformer used to operate from a telephone transmission line into a high quality program repeater. Impedance ratio is 600 to 15,000 ohms.

voltage winding. If the grid capacitance and conductance conditions are not reproduced faithfully in the measuring circuit, the high-frequency voltage amplification may appear to be much better than the true value.

In addition to more precise transmission measuring circuits, various other special circuits have been developed for measuring transformers, such as modulation, impedance, and crosstalk measuring circuits. The design of these circuits is of necessity a specialized art.

In the foregoing, various types of improvements in communication transformers have been discussed. Wherever applicable, several such improvements have been incorporated in an individual design. The improved performance of transformers as described has been an essential step in the development of the communication art.

References

1. H. D. Arnold and G. W. Elmen. Permalloy—an Alloy of Remarkable Magnetic Properties. *Franklin Inst. Journal*, May 1923, p. 621–32.
2. G. W. Elmen. Magnetic Alloys of Iron, Nickel, and Cobalt. *Franklin Inst. Journal*, May 1929, Vol. 207, p. 595.

3. G. W. Elmen, U. S. Patents: 1,586,883; 1,586,884; 1,620,878; etc.
4. G. W. Elmen, U. S. Patent, 1,768,237.
5. C. R. HANNA. Design of Reactances and Transformers Which Carry Direct Current. *A. I. E. E. Journal*, Vol. 46, February 1927, p. 128–31.
6. J. B. JOHNSON. Thermal Agitation of Electricity in Conductors. *Phys. Rev.*, July 1928, p. 97–109.
7. H. S. BLACK. Stabilized Feedback Amplifiers. *Elec. Engg. (A. I. E. E. Trans.)*, Vol. 53, January 1934, p. 114–20.
8. W. L. CASPER. Telephone Transformers. *A. I. E. E. Journal*, Vol. 43, March 1924, p. 197–209.
9. T. SPOONER. Audio Frequency Transformer Characteristics. *Elec. Journal*, July 1926, p. 367–75.
10. F. E. FIELD. The Evolution of the Input Transformer. *Bell Lab. Record*, Vol. 3, October 1926.
11. E. L. SCHWARTZ. Permalloy in Audio Transformers. *Bell Lab. Record*, Vol. 6, April 1928.
12. K. S. JOHNSON. Transmission Circuits for Telephonic Communication (a book). D. Van Nostrand Company, Inc., 1927.
13. T. E. SHEA. Transmission Networks and Wave Filters (a book). D. Van Nostrand Company, Inc., 1929.
14. C. E. LANE. Phase Distortion in Telephone Apparatus. *Bell Sys. Tech. Jour.*, v. 9, July 1930, p. 493–521.
15. EUGENE PETERSON. Harmonic Production in Ferromagnetic Materials. *Bell Sys. Tech. Jour.*, v. 7, Oct. 1928, p. 762–96.

Measurement of Telephone Noise and Power Wave Shape *

By J. M. BARSTOW, P. W. BLYE and H. E. KENT

IN studies of the inductive coordination of power and telephone systems from the noise standpoint, a knowledge of the magnitudes of the harmonic currents and voltages on the power circuits and of the harmonic components of the telephone circuit noise is necessary. It is also necessary that there be available a means of rating and summing up these individual components to give an overall indication of their effects on a person using a telephone connected to one of the exposed circuits. This paper discusses methods which have been developed for making such overall measurements.

The effect on a listener of a given amount of noise on a telephone circuit is a complex one, and it is not practicable in the day-by-day maintenance of telephone circuits to measure separately all the factors involved. Rather, it is necessary to make some overall measurement of the circuit noise which may be related to its effect on telephone transmission. It is, of course, desirable that the measuring devices used should measure different circuit noises as equal when they produce equal interfering effects on telephone transmission.

Two methods of measuring telephone circuit noise are at present in use in the Bell System. One of these methods is subjective, that is, uses the human hearing mechanism as a part of the measuring apparatus. This method consists of comparing, in a telephone receiver, the noise to be measured with a noise generated by means of a standard buzzer. The observer adjusts the magnitude of the buzzer noise by means of a calibrated potentiometer until, in his judgment, it is as disturbing as the noise to be measured.

The objective method of noise measurement which has been made available within the last few years employs an electrical network for weighting the various single frequency components of a noise as closely as practicable in accordance with their interfering effects on telephone transmission, and a calibrated amplifier to raise the energy level of the weighted components sufficiently to operate an electric meter. The chief operating advantages of the objective method are the repro-

* Digest of a paper published in the December 1935 issue of *Electrical Engineering* and scheduled for presentation at the A.I.E.E. Winter Convention, New York, N. Y., January 28–31, 1936.

ducibility of the results and the ease and speed of making measurements. Its disadvantage lies in the difficulty in determining the complex nature of the human hearing mechanism and simulating its characteristics sufficiently well in objective apparatus.

One of the important steps in the development of the objective noise meters has been the determination of the relative interfering effects of different single frequency tones. Two types of tests have been used for this purpose, (*a*) judgment tests and (*b*) articulation tests.

Judgment tests usually are set up so that the observer may compare directly two noises in the presence of speech heard over a representative telephone circuit. The magnitude of one of the noises is adjusted until it is judged to be as disturbing as the second noise. The magnitudes which the observer judges to be equally disturbing can be measured and in the case of single frequency tones, the relative weighting which should be applied to the two frequencies may thus be determined.

An articulation test consists essentially in calling a number of meaningless monosyllables over a circuit to a group of observers, each of whom records the sounds that he hears. The percentage of sounds correctly received is termed the "per cent articulation" for the particular condition tested. On a given circuit, two different noises which produce the same loss in articulation would usually be considered as equally interfering. As before, in the case of two different single frequency noises, measurements may be used to determine the relative weightings to be applied to the two frequencies.

In 1919, the results of judgment and articulation tests on the relative interfering effects of different single frequency tones were published in a paper by H. S. Osborne.[1] Since that time, several other sets of tests of this character have been made in order to check the values previously obtained and to extend the frequency range covered. From the results of all these tests[2] and a recognition of the trend toward more uniform frequency response in telephone message channels, a single curve of relative interfering effects of different single-frequency tones in a telephone receiver has been derived. This is shown in Fig. 1, the curve being labeled " receiver currents." By combining with this curve the frequency characteristic of a representative transmission path between the toll circuit terminals and the

[1] "Review of Work of the Subcommittee on Wave Shape Standard of the Standards Committee," H. S. Osborne, *A. I. E. E. Transactions*, Vol. 38, Part 1, 1919.

[2] The articulation and judgment tests mentioned here also contributed largely to the selection, by the C. C. I. F. (international advisory committee on telephony), of a curve of relative interfering effects of single-frequency tones expressed in terms of voltage across the receiver, which it has recommended as a basis for noise measurement on international circuits. The weighting given in curve A of Fig. 1, when expressed in similar terms is in conformity with the weighting recommended by the C. C. I. F.

telephone subscriber's receiver, a second curve indicating the relative interfering effects of different single-frequency currents in the telephone line has been derived. This is also shown in Fig. 1. These two curves have been incorporated in the indicating noise meters for use in measuring noise in the subscriber's receiver and noise at the terminals of a toll circuit, respectively.

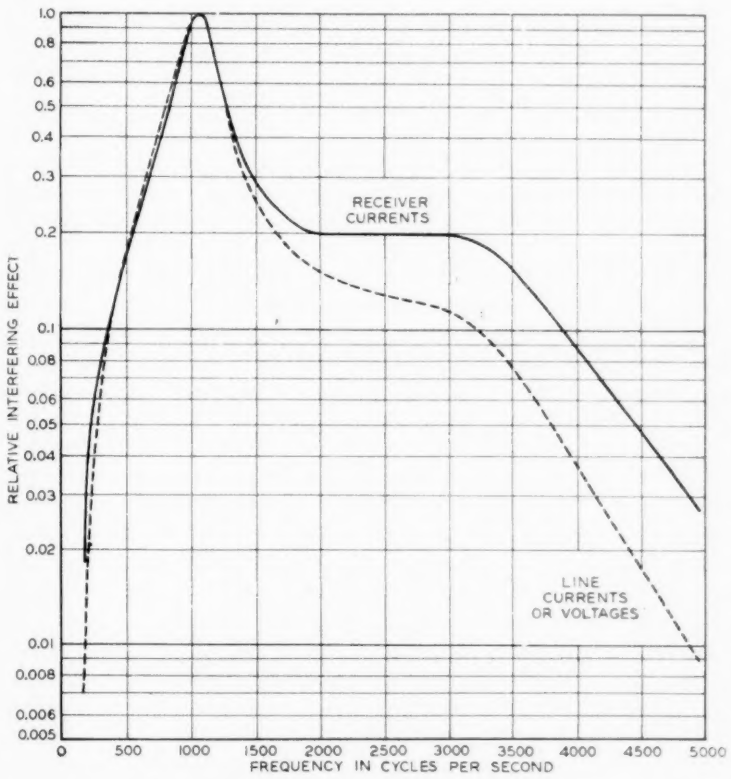


Fig. 1—Relative interfering effects of telephone circuit noise currents.

A second important factor considered was the manner in which various single-frequency noises combine in the human ear. The rule of combination adopted was that by which each single frequency contributed to the total meter reading in proportion to its weighted power. (This is the equivalent of the familiar root-sum-square rule for summing up currents or voltages,)

In addition to the requirements for weighting and rule of combination, it was thought desirable to employ an indicating instrument in which the change of reading was about as rapid as the change in appreciation of loudness in human hearing. From published results and confirming tests it was determined that on the average, the indicating instrument should reach a full deflection for sounds lasting .2 second or longer.

Under these general specifications, several models of circuit noise meters were built and two series of tests were made to determine their adequacy for measuring circuit noise. These tests were made under the auspices of the Joint Subcommittee on Development and Research of the Edison Electric Institute and the Bell System. The first was a rather extensive series of articulation tests on open-wire toll circuit noise. Since none of the toll circuit noises tested contained components of importance above 2,000 cycles, a series of judgment tests was carried out on representative noise of the type arising by induction in telephone circuits exposed to a-c. lines supplying rectifiers and on various high-frequency noises derived therefrom.

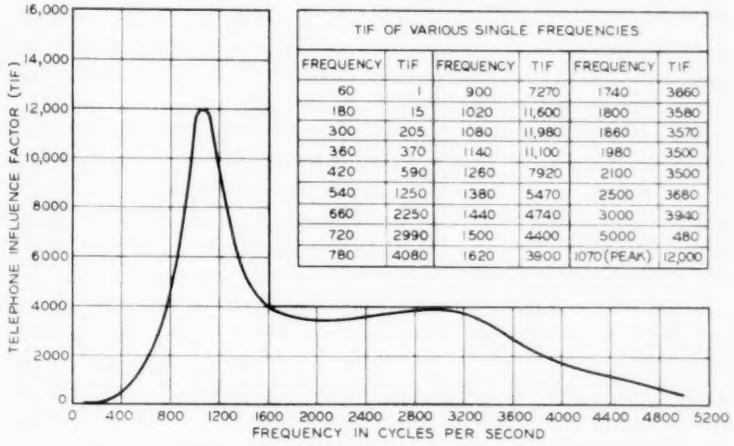
The articulation tests showed that when toll circuit noises of various types produced equal losses in articulation under the given set of telephone conditions, they were measured as substantially equal by both the objective and subjective methods of measuring. The objective method gave a slightly better correlation than did the subjective method even though the average of 18 individual observers was used in the latter. While the correlations were not as close with the high-frequency noises as in the case of the more common types of toll circuit noise, on the average the noise meter rated the rectifier noises at least as well as did the ear balance method, the latter using 10 observers.

A device called the "Telephone Interference Factor Meter" for measuring or rating the wave shape of power system currents and voltages in terms of their influence on exposed telephone circuits was described in the Osborne paper of 1919, referred to above. With this instrument, an indication was obtained of the total harmonic content of a given voltage wave, the individual components present being weighted approximately in proportion to their relative interfering effects.

The data obtained from the more recent studies of relative interfering effects described above have made possible a revision of the method of measuring T.I.F., in which the basic principle has been retained but in which the frequency weighting characteristic has been revised somewhat and its range extended to about 5,000 cycles. In

connection with this revision, the name has been changed to "Telephone Influence Factor."

The Telephone Influence Factor (T.I.F.) of a voltage or current wave is the ratio of the square root of the sum of the squares of the weighted effective values of all the sine wave components (including, in alternating current waves, both fundamental and harmonics) to the effective value of the wave. The weightings decided upon to be applied to the individual components are as shown in Fig. 2.



TIF OF VARIOUS SINGLE FREQUENCIES

FREQUENCY	TIF	FREQUENCY	TIF	FREQUENCY	TIF
60	1	900	7270	1740	3660
180	15	1020	11,600	1800	3580
300	205	1080	11,980	1860	3570
360	370	1140	11,100	1980	3500
420	590	1260	7920	2100	3500
540	1250	1380	5470	2500	3680
660	2250	1440	4740	3000	3940
720	2990	1500	4400	5000	480
780	4080	1620	3900	1070 (PEAK)	12,000

Fig. 2—Frequency weighting characteristic for TIF measurements.

In deriving the revised frequency weighting characteristic, the following factors representing distortion occurring in the various media intervening between the power circuit current or voltage and the telephone subscriber's ear were considered.

1. Relative interfering effects of single-frequency components in the receiver of a subscriber's telephone set.
2. Distortion occurring between the terminals of the circuit in which the noise is induced and the subscriber's receiver.
3. Variation in coupling between power and telephone circuits with frequency.
4. Variation of effects of telephone circuit unbalances with frequency.

Data on Items 1 and 2 above were combined to derive the line weighting characteristic of the telephone circuit noise meters indicated by the "line currents" curve of Fig. 1. It was, therefore, possible to use this curve directly to represent the combination of these two

factors. A factor directly proportional to frequency was adopted to represent inductive coupling between power and telephone circuits (Item 3), the work of the Joint Subcommittee on Development and Research having indicated that, in general, coupling may be so represented. After studying data available on Item 4, it was concluded that no type of frequency weighting could be adopted which would satisfactorily represent all types of telephone circuit unbalances. Thus T.I.F. as measured by the method described here is a correct index to the influence of a power circuit voltage or current only for those cases where unbalances are independent of frequency. This is usually the case on open-wire toll circuits and open-wire exchange circuits employing bridged ringers. In other cases some empirical modification may be necessary.

Since a large amount of data has been obtained with the old T.I.F. meter, it was considered desirable, if practicable, to adjust the scale of the revised set so that readings made by this method would, in general, be approximately the same as readings obtained by the old meter. In this connection calculations using the old and new weightings were made on a large number of machines and circuits of various types on which harmonic analyses were available. These calculations were supplemented by a considerable number of comparative measurements in the factory and in the field, using meters employing the old and new weightings. These calculations and tests indicated that in the average case, reasonably satisfactory correlation between the readings made by the two meters would result if a peak value of 12,000 were assigned to the new weighting characteristic, as shown in Fig. 2.

Several experimental models of T.I.F. measuring sets were made employing the new weighting characteristic and these have given very satisfactory results. The adoption of the rule that coupling is to be considered proportional to frequency also makes it possible to use a circuit noise meter and a small amount of auxiliary apparatus to form a T.I.F. meter.

The development of the revised method of measuring T.I.F. has also been conducted under the auspices of the Joint Subcommittee on Development and Research of the Edison Electric Institute and the Bell Telephone System.

On the Correlation of Radio Transmission with Solar Phenomena *

By A. M. SKELLETT

A DAILY character figure for radio transmission is obtained from the data of the short-wave transatlantic telephone circuits of the American Telephone and Telegraph Company. The New York-London circuits are in practically continual use so that they furnish data from which a character figure, representative of the whole 24 hours, may be derived. Such figures are based on the ratio of uncommercial to total time and thus are indirectly dependent on field strengths.

In order to facilitate plotting, these character figures were reduced to 3 group indices. Figure 1 shows the indices arranged to bring out the twenty-seven-day recurrence tendency. This is demonstrated by the apparent bunching of the spots into more or less vertical columns. Terrestrial magnetic data are shown alongside in similar form for comparison.

The recurrence tendency is well enough marked in the chart so that useful predictions of future behavior may be made. The chart is kept up to date and then by inspection a prediction may be made for any day not more than twenty-seven days distant. Some idea of the probable accuracy may also be obtained from the chart by noting whether the day in question falls, for instance, in the middle of a major sequence or on the ragged edge of a poorly defined one. Such probable accuracy is expressed by modification of the prediction with the words "probably" or "possibly."

The correlation between the two phenomena is good enough so that predictions of activity of one nature may be made from the chart of the other type of activity. For instance it would be possible to predict the radio behavior from the magnetic chart alone. This method has been found to yield the same order of accuracy as that using the radio chart alone.

Daily predictions of the behavior of the radio circuits from either the radio or magnetic chart have been correct 62 per cent of the time. Similar predictions of the magnetic data from the magnetic chart have

* Digest of a paper presented at the Sixteenth Annual Meeting of the American Geophysical Union, Washington, D. C., April 25, 1935, and published in full in the *Proceedings of the Institute of Radio Engineers*, November, 1935.

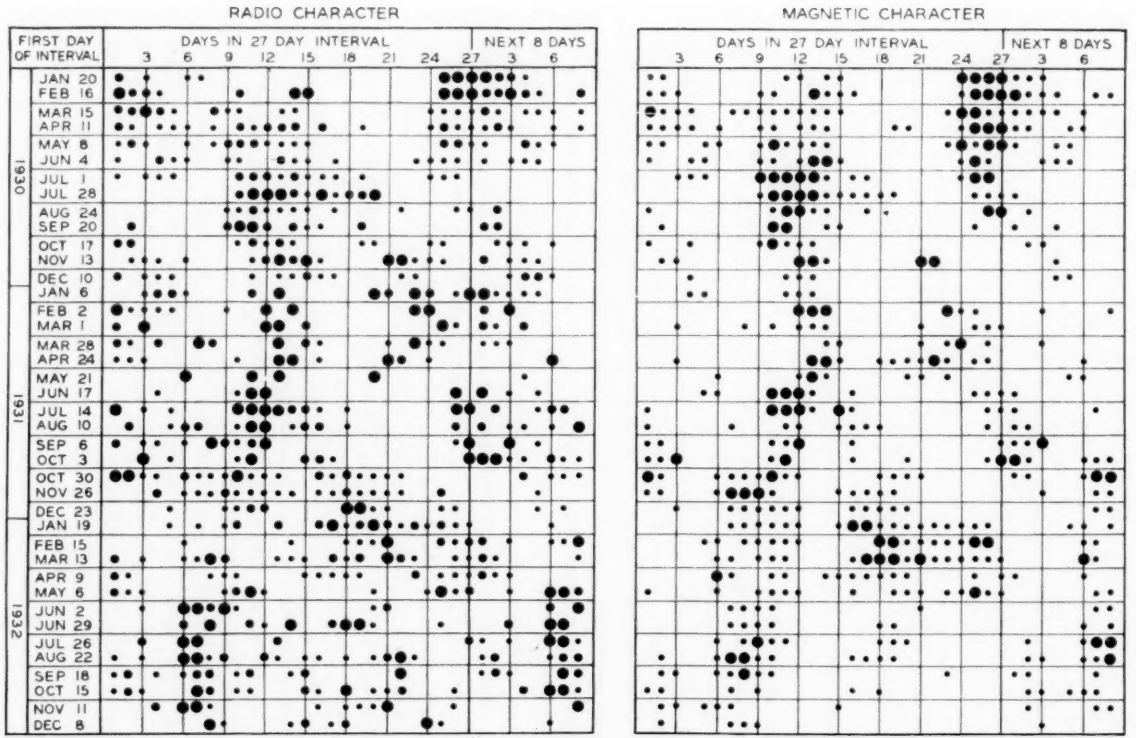


Fig. 1—Relative day-to-day record of short-wave radio transmission over the North Atlantic Ocean and terrestrial magnetism activity for 1930, 1931 and 1932, demonstrating the 27-day recurrence tendency. The dot size corresponds to the severity of disturbance.

been correct 71 per cent of the time. These figures have been determined solely on the basis of "disturbed" or "undisturbed" and modification of the forecasts by the words "probably" and "possibly" have not been taken into account.

This method of making predictions, even in its present state, is of definite use commercially. Special forecasts of the same nature have proved useful in planning certain experimental studies.

Well defined sequences of activity of an approximately 27-day period are also apparent in data of solar phenomena, particularly those relating to sunspots, bright and dark hydrogen flocculi,[1] and prominences. An attempt to link such solar sequences with the terrestrial ones noted above in a cause and effect relationship was not, however, very successful. For several well marked radio (and magnetic) sequences it was found that no single type of solar activity could be identified in such a manner as to exhibit a clear cut relationship. For some of the sequences there was some form of solar activity near the center of the sun at the time of each radio disturbance but such activity varied between recurrences in heliographic latitude and longitude and in kind.

A similar indefinite result was found by starting with the solar sequences and attempting to match the terrestrial data with them. For instance an area on the sun approximately in heliographic latitude + 10° and longitude of 315° to 329° exhibited the presence of either hydrogen flocculi, prominences, or sunspots or combinations of these on each transit across the face of the sun from October 21, 1932 to February 9, 1933, a total of five transits. Sunspots appeared in this region on the last four transits and their identity over this period of time was noted at Mt. Wilson Observatory.[2] Although the times of central meridian crossing of this area fall within a well defined sequence[3] on the radio chart (between days 6 and 9 on the left at the bottom of the chart) the absence of activity on this solar area for earlier recurrences of the radio sequences tends to vitiate the relationship between radio disturbances and those types of solar activity which were observed. Nevertheless, the reality of the 27-day period in radio is strong indication that solar activity is responsible, even though not convincingly identified in detail.

Various other criteria were used for segregating the solar data for correlation. For instance, a study of the solar distribution of flocculi

[1] Hydrogen flocculi are clouds of hydrogen gas observed with a spectrohelioscope set on the hydrogen line H_α.

[2] *Publ. Astr. Soc. Pac.*, 45, 263, 53, Feb. 1933.

[3] This sequence is considerably strengthened by increasing the number of group ndices into which the data are divided.

and spots on terrestrially disturbed and quiet days shows a maximum and minimum, respectively, about 13° west of the center (one day past); see Fig. 2. These curves are interpreted as indicating that the

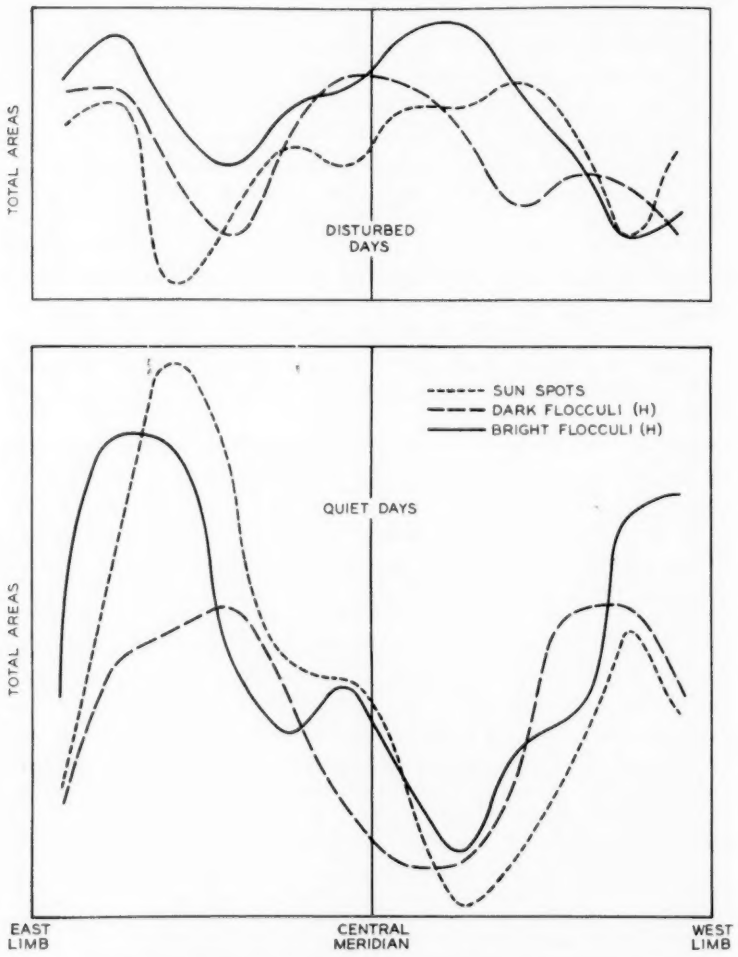


Fig. 2—Distribution across the solar disc of solar phenomena on terrestrially disturbed and quiet days.

most probable position of flocculi and spots on disturbed days is 13° west of the center of the solar disc and that on quiet days it is other

than in this region. The one-day interval from the center is interpreted as the time taken for the propagation of the disturbance from the sun to the earth.

Probably one reason for the indecisive nature of the results is to be found in the intermittent manner in which the solar data are necessarily obtained. It seems likely that considerably more success might be obtained in determining the solar-terrestrial relationships, if the solar disc could be watched continually on a world-wide program of observation, as Hale [4] has suggested, to record all solar outbursts and so to increase the completeness of the solar data until they approach those of the terrestrial.

[4] *Astrophys. Jour.*, 73, 408, 1931.

Eclipse Effects in the Ionosphere *

By J. P. SCHAFER and W. M. GOODALL

It is concluded from measurements of virtual heights and critical ionization frequencies of the various regions of the ionosphere which were made during two solar eclipses at Deal, New Jersey, that ultra-violet light is an important ionizing agency in the E, M, F_1, and F_2 regions.

AS a result of pulse measurements made at Deal, New Jersey, during the partial eclipse of the sun February 3, 1935,[1] and during the total eclipse of the sun of August 31, 1932,[2] we now have data which show that the passage of the moon's shadow across the earth is accompanied by a decrease in ionization in four of the ionized regions of the ionosphere (E, M, F_1 and F_2).[3]

During the 1932 eclipse the ionic density in the E and F_1 regions was found to decrease, with the maximum effect occurring shortly after the eclipse maximum. Since the ionization in these two regions ordinarily changes uniformly with time, and since the variations observed during the eclipse were much larger than normal variations, we believe that the decrease in ionic density was actually caused by the eclipse. As regards the changes observed in the F_2 region, our 1932 results were not conclusive because the maximum effect in this region did not coincide with the eclipse but occurred somewhat later. The ionic density in this region is known to fluctuate on at least some non-eclipse days and did in fact undergo comparable variations on several occasions during the two days preceding and the two days following the eclipse. Other observers have reached the same conclusions as regards the F_2 region during the 1932 eclipse from their own data.[4]

The data from which our conclusions were drawn are shown in Fig. 1.

* Presented before joint meeting of I. R. E. and U. R. S. I., Washington, D. C., April 26, 1935. Published in same brief form in November, 1935 *I. R. E. Proceedings* as in this *Journal*.

[1] Letter to *Nature*, vol. 135, p. 393; March 9, 1935.

[2] Mention has already been made of the results of our 1932 eclipse experiments in the following publications:

Science, November 11, 1932; *Proc. Fifth Pacific Science Congress*, vol. 3, pp. 2171–2179, 1934; *Nature*, September 30, 1933; *Bell Lab. Record*, March, 1935.

The data have never been published and we are therefore including some of it in this paper as it may be of interest to other investigators in this field.

[3] M refers to the intermediate region between E and F_1.

[4] Kirby, Berkner, Gilliland, and Norton, *Proc. I. R. E.*, vol. 22, pp. 246–265, February, 1934; Henderson, *Canadian Jour. Res.*, January, 1933.

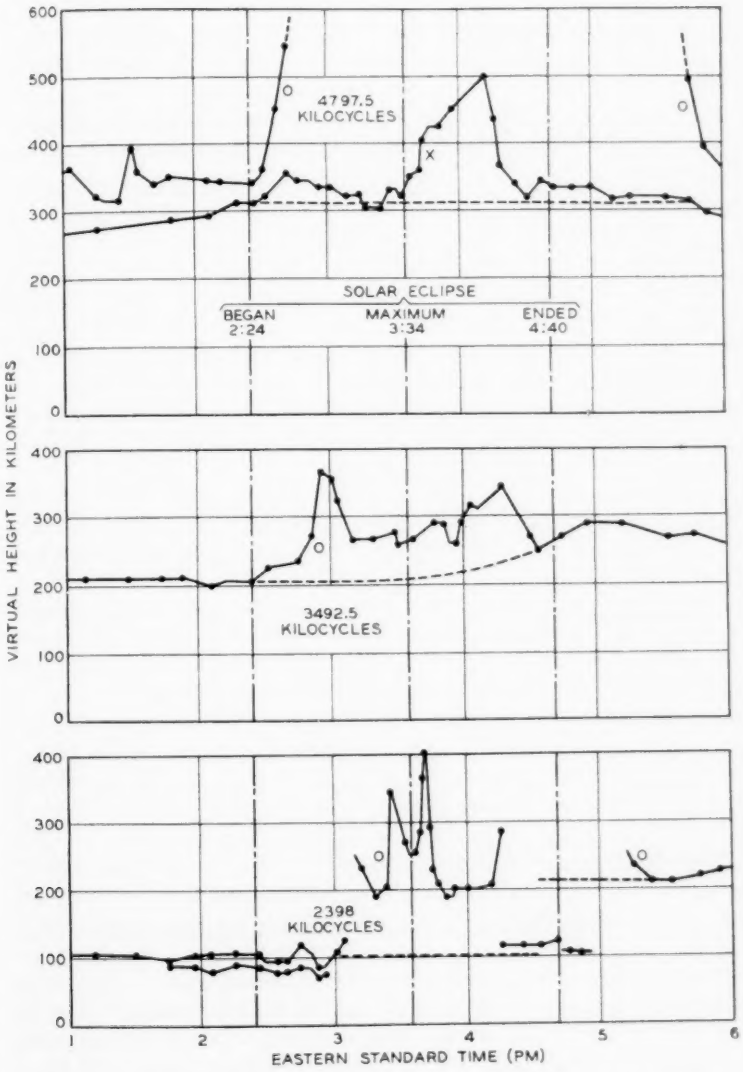


Fig. 1—Virtual height values during the total eclipse of August 31, 1932.

The double maximum in virtual height with a minimum between for 2398 and 3492.5 kilocycles [5] is interpreted by us to have been caused by a decrease in ionic density in the F_1 region, which resulted in a change in reflection from the F_1 to the F_2 layer during the central part of the eclipse.

From the curves of Fig. 1, it is possible to plot the virtual height contour map shown in Fig. 2. Since we know as a result of data taken

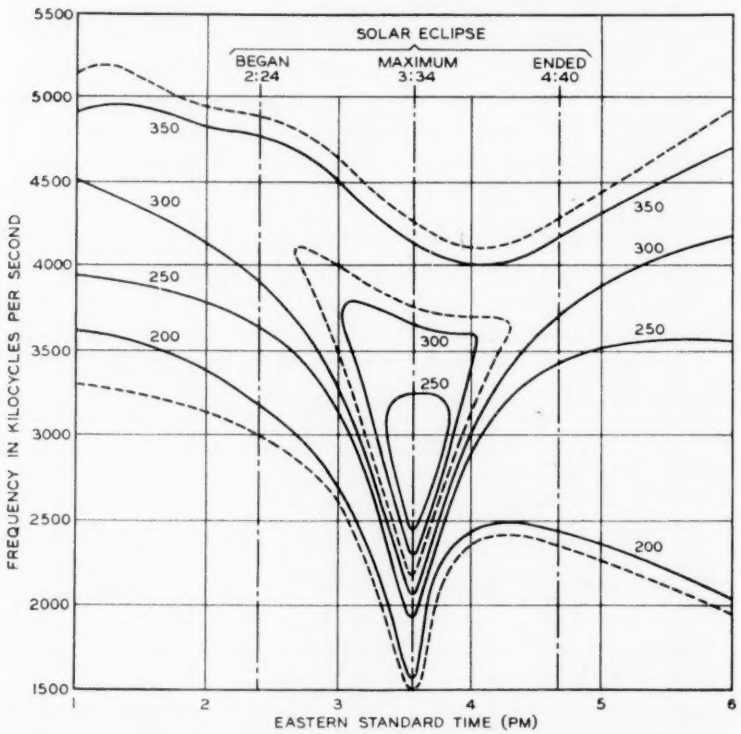


Fig. 2—Virtual height contour map drawn from data for Fig. 1, for August 31, 1932.

on several hundred days that the *X*-component curve in Fig. 1 for 4797.5 kilocycles is nearly equivalent to the *O*-component [6] curve for a frequency approximately 750 kilocycles lower (i.e., 4050 kilocycles),

[5] Mimno and Wang, *Proc. I. R. E.*, vol. 21, pp. 529–545, April, 1933, and Kenrick and Pickard, *Proc. I. R. E.*, vol. 21, pp. 546–566, April, 1933, obtained similar double maximum curves.

[6] The expressions "*O*-component" and "*X*-component" are used in place of the terms "ordinary ray" and "extraordinary ray" used by other writers.

there are in effect *O*-component curves for four different frequencies available for plotting the contour map. The dotted lines represent regions of maximum ionization. These curves show in a rather striking manner the sharp decrease in ionization of the E and F_1 regions near the time of the eclipse maximum.

The uncertainty of the 1932 results as regards the F_2 region led us to concentrate our efforts on this region during the 1935 eclipse. Improved technique now made it possible to measure the critical ionization frequencies directly instead of making virtual height measurements on fixed frequencies as had been done during the 1932 eclipse. The critical frequencies for the E and M regions were measured in addition to those for the F_2 region.

We found that this eclipse was accompanied by a decrease in the maximum ionic density in all three regions and that the minimum ionization occurred at or very shortly after the eclipse maximum. The percentage decrease was progressively greater from the lowest to the highest region, being approximately twenty per cent for the E region, twenty-two per cent for the M region, and twenty-five per cent for the F_2 region.

Some such progressive change might be expected from the fact that the eclipse had a magnitude of forty per cent at the ground and approximately forty-three per cent in the F_2 region (250 kilometers over Deal). These magnitudes are in terms of the sun's diameter, and on the basis of eclipsed area these figures become twenty-nine and thirty-one per cent, respectively.

Figure 3 gives the critical ionization frequencies for the three days on which data were obtained. The curves for the E and M regions are for the *O*-component while the curves for the F_2 region are for the *X*-component. For the *O*-component, the ionic density, N, is proportional to f_c^2, where f_c is the critical ionization frequency, while for the *X*-component curve in Fig. 3, N, is proportional to $(f_c - 750 \text{ kc.})$.[2]

The decreases in ionic density of the various regions may be compared with a fifty to sixty per cent decrease in the E region ionization during the eclipse of August 31, 1932, when the eclipse magnitude was ninety-five to one hundred per cent.

The 1935 measurements give a more definite synchronism than those of 1932, between the eclipse occurrence and the time of decrease in ionic density of the F_2 region.

In view of the variable nature of the F_2 region, it is a possibility that the decrease in ionic density at the time of the eclipse was a mere coincidence and was actually due to some noneclipse agency. We believe that this was not the case and that the decrease in F_2 ionization

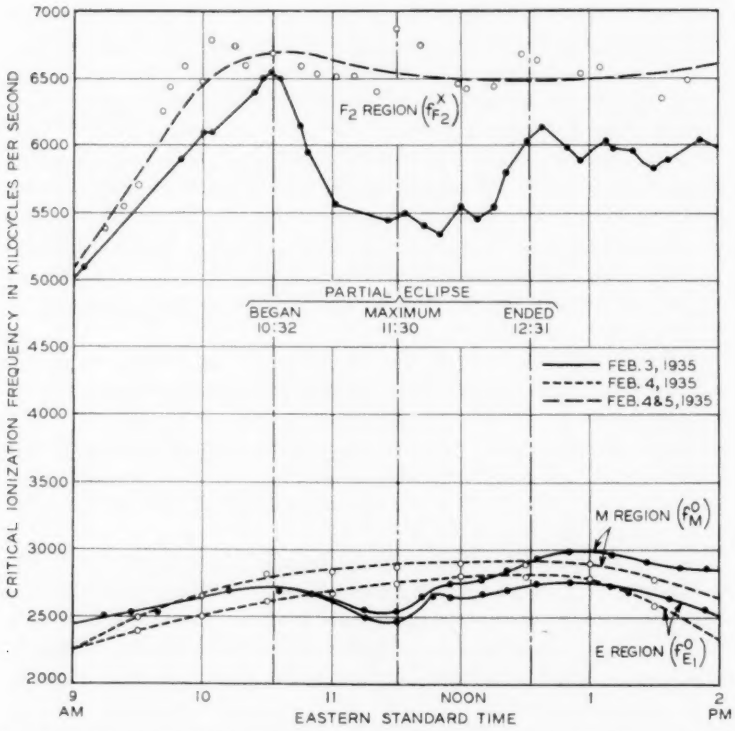


Fig. 3—Critical ionization frequencies during the partial eclipse of February 3, 1935

was a bona fide eclipse effect, as the decrease began within a few minutes after the first contact, and the density attained its lowest value shortly after the maximum of the eclipse and recovered to a more or less constant higher value a few minutes after the last contact. At no time during these measurements on the eclipse day or the days after was there any other variation of a comparable magnitude.

These results therefore indicate that ultraviolet light [7] is an important ionizing agency in the E, M, F_1 and F_2 regions of the ionosphere.

[7] While ultraviolet light is probably the ionizing agency responsible for the effects noted, any other solar emanation which travels substantially at the velocity of light should not be precluded from consideration. *See* E. A. W. Müller, *Nature*, February 2, 1935, who suggests Roentgen type radiation.

Earth Resistivity and Geological Structure *

By R. H. CARD

IN connection with inductive coordination problems it is frequently necessary to estimate low-frequency ground-return mutual impedances between power and communication lines. The distribution of currents in the earth is a major factor in the determination of these impedances. This distribution is controlled by the resistivities of the component parts of the earth's crust and the arrangement of these parts. In impedance formulas that are customarily used the effect of the earth is taken care of by the inclusion of a single parameter—the earth resistivity. For a homogeneous earth this would be the actual resistivity of the material composing it. But the crust is nowhere homogeneous; hence, the resistivity used in such formulas is always of the nature of an average of the resistivities of the several parts of the crust—it is termed the effective earth resistivity.

The effective earth resistivities for fundamental power-system frequencies derived from mutual impedance measurements made in many parts of the world range, in general, from 2 to 10,000 meter-ohms. In a few instances values considerably higher than 10,000 meter-ohms have been observed. (The resistivity of a particular material, expressed in terms of the meter-ohm, is equal to the resistance in ohms between opposite faces of a one-meter cube of that material.)

With such a range of earth resistivities to contend with it is to be expected that estimates of ground-return mutual impedances for situations in areas where no earth resistivity data are available may be in error by large factors. In an effort to improve upon the accuracy of such estimates a study was begun several years ago of the relation between effective earth resistivity and geology. Consideration was at that time given only to areal geology, the geology of the strata of the crust lying immediately below the soil and other loose surface materials.

From this preliminary work, it appeared that the resistivities in areas of very old rocks were high and that, in a general way, decreasing resistivity corresponded to decreasing age of the rocks. There were, however, a number of outstanding discrepancies that could not be satisfactorily explained.

* Digest of a paper published in *Electrical Engineering*, November, 1935, and scheduled for presentation at the A. I. E. E. Winter Convention, New York, N. Y., January 28–31, 1936.

It then became apparent that consideration of the areal geology alone was not sufficient; instead, that the earth's structure to depths ranging from several hundred to several thousand feet must be taken into account. Data on this structure and the effective resistivities indicated by mutual impedance measurements at a large number of test sites have now been assembled. Analysis of these data shows a more or less consistent relation between the resistivity at any given point and the age and physical characteristics of the geological formations involved. This relation is such that, in general, decreasing effective resistivity corresponds roughly to decreasing age of the formations, as the earlier study seemed to indicate. However, there are certain exceptions to this rule.

The principal correlation data derived from the tests are summarized in the following tabulation. This is the result of grouping the tests in accordance with the geological periods to which the principal strata comprising the structure in each case belong and noting the ranges within which the resistivities determined by the greater part of the tests of each group lie.

Pre-Cambrian and combinations of Pre-Cambrian and Cambrian	1000–10,000 m.-ohms
Cambrian and Ordovician combinations	100–1000 m.-ohms
Ordovician to Devonian, inclusive, and combinations of these periods	50–600 m.-ohms
Carboniferous, Triassic, and combinations of Carboniferous and earlier periods	10–300 m.-ohms
Cretaceous, Tertiary, Quaternary and combinations of these periods	2–30 m.-ohms

It would be well to examine briefly the meaning of this summary and to consider its limitations. The geologists tell us that underlying the entire continent are extremely old rocks, extending downward to great depths. Little is known of the relative ages of different parts of this underlying structure. They are here grouped under the general term pre-Cambrian. In some areas pre-Cambrian rocks appear at or near the surface, the only covering being clays, soils, and other loose materials. In other areas they are overlain by rocks and sediments formed during later periods, the total thickness of which ranges up to many thousands of feet. The ages, arrangement, and characteristics of these upper strata are much better known. They are assigned by geologists to various periods in accordance with the ages during which they were formed. These periods appear in the tabulation in order from the oldest to the youngest.

In the case of tests made in areas where the pre-Cambrian rocks are overlain by younger strata it becomes necessary to consider just what portion of the structure probably influenced the test results. For

instance, at the test sites included in the second group of the summary the upper part of the structure consists of Ordovician rocks. These are underlain by Cambrian strata which, in turn, lie on the pre-Cambrian base. The question arises whether the results were influenced by the Ordovician strata alone, by both the Ordovician and Cambrian, or by the Ordovician, Cambrian, and pre-Cambrian. Calculations have been made which indicate that probably only the Ordovician and Cambrian strata were involved to any important extent. The tests in the other groups have been treated in a similar manner.

In the cases considered above the measurements were apparently influenced largely by strata of a single period or of two or more periods of about the same age. The problem is not always as simple as this. For instance, in some areas combinations of very old and very young strata occur. Areas in which the oldest rocks, the pre-Cambrian, lie directly under comparatively thin sediments of the latest periods—the Quaternary, Tertiary, and Cretaceous—are not uncommon. The effective resistivities shown by the tests in such areas range between very wide limits and the tabulation should not be taken as indicative of the values which may prevail under such conditions.

The effects of soils, glacial drift, alluvial deposits along the courses of streams, and other surface materials may also in some instances be such as to result in effective resistivities differing widely from those that would be indicated by the tabulation. The effect of local alluvial deposits, where they overlie the older rock strata, is to lower very materially the effective resistivity that would be expected were the deposits not present.

Another limitation which must be considered is concerned with the presence of rocks formed by volcanic action. Apparently such rocks usually have a high resistivity and where they occur in a comparatively young structure, the effective resistivity may be much higher than the tabulation would indicate.

The test results indicate also that the effective resistivities of structures of given periods within certain large geographical regions are markedly different from those of structures of the same periods in other large regions. Within any one of such regions, excluding areas where igneous and highly metamorphosed sedimentary rocks are involved, the effective resistivities of structures of the same period are encompassed within a comparatively narrow band.

To facilitate the use of the correlation data, the different areas within which tests have been made have been divided into groups in accordance with the geological periods of the uppermost strata within these

areas and maps have been prepared which show the boundaries of these areas and the results of the tests within them. Two such maps are shown in Figs. 1 and 2. These maps show the geological periods of the upper strata as they would appear were the overlying mantle of soil, glacial drift, local alluvial deposits and such removed. Any one

Fig. 1—Areal geology and effective resistivities—pre-Cambrian areas.

"test," as the designation is employed here, may include measurements of ground-return mutual impedances for a number of different conductors or sections of line and the test results for these different conductors or sections may indicate quite a wide range of resistivities—often 2 or 3 to 1 and occasionally 10 or more to 1, depending upon the

degree of complexity of the earth's structure at the test site. The maps show, for each test, roughly the average of the resistivities indicated by the different measurements. They also indicate for each test the relative extent of the lines involved in the test.

Within the heavy boundary lines of Fig. 1 the rocks are largely pre-Cambrian. It will be noted that most of the tests in these areas indicated very high earth resistivities—from 1400 to 10,000 meter-

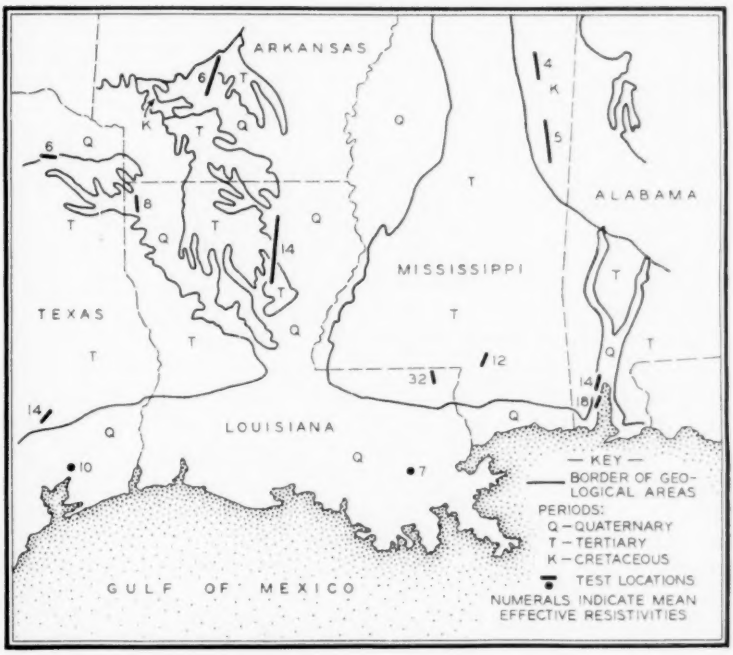


Fig. 2—Areal geology and effective resistivities—Gulf Coastal Plain areas.

ohms. The single exception, one test in New York which showed a value of 200 meter-ohms, is illustrative of the effect of alluvial deposits. The lines involved in this test were located in a narrow valley, the floor of which was partly covered with alluvium. By contrast, the structures of the areas shown in Fig. 2 are of the three latest periods, the Quaternary, Tertiary and Cretaceous, and the effective resistivities are very low—from 4 to 32 meter-ohms.

Abstracts of Technical Articles from Bell System Sources

Heat Treatment of Magnetic Materials in a Magnetic Field—II. Experiments with Two Alloys.[1] RICHARD M. BOZORTH and JOY F. DILLINGER. The magnetization of two alloys, as affected by heat treatment in a magnetic field at various temperatures, is examined in some detail in order to elucidate the nature of the accompanying changes which result in some cases in a 30-fold increase in maximum permeability. The *experiments* show that these alloys (one containing approximately 35 per cent iron and 65 per cent nickel, the other 20 per cent iron, 60 per cent cobalt and 20 per cent nickel) can be effectively heat treated in a magnetic field of 10 oersteds if the temperature is above 400° C. and below the Curie point of the alloy. The time during which the magnetic properties change has been measured at different temperatures and is found to vary according to the equation $\tau = Ae^{W/kT}$. The experiments are *interpreted* in terms of the domain theory of ferromagnetism. The changes which occur are due to the relief of magnetostrictive stresses which arise when the material becomes ferromagnetic upon cooling through the Curie point or when an external magnetic field is applied, and the relief comes about by plastic flow or diffusion in the separate domains. The values of A (about 10^{-12} second) and W (2.1 electron volts) are the same as those determined by Bragg and Williams for the above equation which also gives the time necessary for the establishment of a superstructure in alloys. The relation between the two processes, establishment of superstructure and the relief of magnetostrictive strains, is pointed out.

Heat Treatment of Magnetic Materials in a Magnetic Field—I. Survey of Iron-Cobalt-Nickel Alloys.[2] JOY F. DILLINGER and RICHARD M. BOZORTH. The changes that occur in the magnetic properties of iron-cobalt-nickel alloys when they are annealed in a magnetic field, have been investigated for a series of these alloys. The maximum change for the iron-nickel alloys occurs between 65 and 70 per cent nickel and is evidenced by a large increase in maximum permeability and a hysteresis loop of rectangular shape. All of the alloys with Curie points above 500° C. and with no phase transformation have their properties similarly changed. Thorough preliminary annealing

[1] *Physics*, September, 1935.
[2] *Physics*, September, 1935.

enhances the effect. With an extreme preliminary anneal of 1400° C. for 18 hours specimens of 65 permalloy have been obtained with the record value of maximum permeability of 600,000. The magnetic characteristics of materials treated in this way are relatively insensitive to stress. These magnetic characteristics are, however, highly anisotropic; the maximum permeability in one direction is as much as 150 times as large as that at right angles.

Newer Concepts of the Pitch, the Loudness and the Timbre of Musical Tones.[3] HARVEY FLETCHER. It has generally been thought that corresponding to the three psychological aspects of a sound, namely, the pitch, the loudness and the timbre, there are the three physical aspects of a sound wave, namely, the wave-length, the amplitude and the wave form. Although it is true that there is such a correspondence in a very approximate way, when the matter is examined more closely it is found that each of the psychological aspects depends upon all three of the physical properties of the sound wave.

In the paper it was shown how loudness can be defined in a quantitative way and measured by experimental methods which are described. From such measurements a relation has been found between loudness as it is ordinarily understood by the lay man and the physical intensity. In the higher intensity regions it is found that if the intensity of a sound is increased 1000-fold then the loudness will be increased 10-fold. In other words in these regions the loudness as determined by the average observer is proportional to the cube root of the intensity. For the lower intensities the loudness increases more rapidly than for the high intensities, being almost proportional to the intensity in the regions near the threshold. It is shown that the loudness depends upon the frequency, the overtone structure and the intensity of the complex sound.

In a similar way a precise definition of pitch is given which makes it possible to make quantitative measurements of this psychological aspect of a sound. Contrary to the usual notion it is found that the pitch varies not only with the fundamental frequency but also with intensity of the sound and with the overtone structure. For example, it was found that the pitch of a tone having a frequency between 100 and 200 cycles may be lowered more than a full tone by increasing the intensity without changing the frequency. Also it was shown that the pitch of a complex tone may shift as much as 1 or 2 octaves by changing the overtone structure. Numerous examples are given to show that pitch also depends upon the three physical quantities, frequency, overtone structure, and intensity. Although quantitative measurements

[3] *Jour. Franklin Institute*, October, 1935.

on timbre are still lacking there is no doubt that similar results will be found for timbre, namely, that although it depends principally upon the overtone structure, nevertheless, changes in fundamental frequency and changes in the intensity also produce large changes in the timbre.

Ceramics in the Telephone.[4] A. G. JOHNSON and L. I. SHAW. In this descriptive paper by Western Electric engineers the problems of ceramic materials as they relate to telephone usage are discussed. New products with specific properties include every type of ceramic material—electrical porcelain, vitreous enameled parts, glass and heavy clay products. Manufacturing problems of the above materials are discussed.

La Transformation Triangle—Étoile pour des Éléments de Circuits Généraux.[5] JOHN RIORDAN. This paper gives the relations between the constants of linear passive transducers connected in star and delta. The delta-star transformation previously given by Lavanchy (*Revue Générale de l'Électricité*, XXXVI, pp. 11–31 and 51–59) is shown to admit a slight generalization, perhaps of little practical importance. In the reverse (star-delta) transformation (not previously given), three of the nine independent constants of the three transducers are shown to be defined uniquely, but of the remaining six, three may be defined at pleasure, subject only to dimensional and cyclic requirements. This lack of uniqueness is shown concordant with the connection conditions.

Flutter in Sound Records.[6] T. E. SHEA, W. A. MACNAIR, and V. SUBRIZI. Frequency modulation of a sound signal is caused by non-uniformity in the record speed during the recording or reproducing process. This source of flutter is discussed and was demonstrated at the May 1935 Convention of the Society of Motion Picture Engineers. The paper includes a discussion of the physical nature of frequency modulation, the physiological effects of frequency modulation, the methods of producing known amounts of artificial flutter, and the methods of measuring flutter.

Acoustic Impedance of Small Orifices.[7] L. J. SIVIAN. Data are presented giving the measured acoustic reactance and resistance for a number of circular orifices varying in diameter from 1 cm. down to 0.034 cm., and for a rectangular orifice 1.9 cm. × 0.075 cm. The measurements were made for various particle velocities, the cor-

[4] *Indus. and Engg. Chemistry*, November, 1935.
[5] *Revue Générale de l'Électricité*, September 21, 1935.
[6] *Jour. S. M. P. E.*, November, 1935.
[7] *Jour. Acous. Soc. Amer.*, October, 1935.

responding Reynolds numbers varying from 0.7 to 3000, roughly. The reactance is found substantially independent of the particle velocity; a formula for computing it is given. The resistance approaches a constant value as the velocity is sufficiently decreased; formulae for computing this "low velocity" resistance are given. At larger velocities the resistance increases with the velocity. This is discussed from the standpoint of a loss of kinetic energy of flow, acting besides viscosity and turbulence.

Earth-Potential Measurements Made During the International Polar Year.[8] G. C. SOUTHWORTH. Data are presented covering the normal diurnal variation of earth-potentials as measured at about a dozen different points, mostly in eastern United States. These data are arranged in graphical form for the convenience of the casual reader and also in numerical form for the use of the correlator. The data for Wyanet (Illinois), Houlton (Maine), and New York (New York) are based on nearly continuous recordings extending over a period of one or two years. This period includes the International Polar Year. At other points, less extensive data were taken. These show the general characteristics peculiar to the location in question.

The data taken at Wyanet, Houlton, and New York have been analyzed for harmonic content. At New York the fundamental and to a large extent the harmonics also, are directed along a northwest-southeast line. At Wyanet and Houlton these components tend to rotate with time. The pronounced directive effect noted near New York appears to prevail rather generally along the eastern part of the United States from Massachusetts to Florida and possibly into Cuba. The rotary effect noted in the Houlton and Wyanet data is also found in data taken in the southern part of the Mississippi Valley. Most of the data point toward the generally accepted view that there is a close relation between earth-resistivity and the direction and magnitude of earth-potentials. However, there are some inconsistencies noted which tend to make this less definite.

The Characteristics of Sound Transmission in Rooms.[9] E. C. WENTE. The characteristics of electrical circuits used for communication purposes are advantageously determined from a measurement of the transmission loss as a function of frequency. Similarly, a measurement of the acoustic pressures at various points in a room while sound of fixed intensity is emitted from a source should permit an evaluation of the acoustic characteristics of the room. Measurements of this

[8] *Terrestrial Magnetism*, September, 1935.
[9] *Jour. Acous. Soc. Amer.*, October, 1935.

type have in the past not led to any useful results because of the large variations in acoustic pressures produced by standing waves. When a high-speed level recorder is used to record pressure levels automatically as the frequency of a constant source is continuously varied, curves are obtained which not only show the variations in the general level of acoustic pressures in the various frequency regions, but which also permit an evaluation of the reverberation characteristics of different parts of the room. The paper shows curves obtained with the recorder under various room conditions.

Contributors to this Issue

AUSTIN BAILEY, A. B., University of Kansas, 1915; Ph.D., Cornell University, 1920; Instructor in Physics, Cornell University, 1915–18; Signal Corps, U. S. A., 1918–19; Assistant Professor of Physics, University of Kansas, 1921–22. American Telephone and Telegraph Company, Department of Development and Research, 1922–34; Bell Telephone Laboratories, 1934–. Dr. Bailey's work has been largely along the line of methods for making radio transmission measurements and of long-wave radio problems.

J. M. BARSTOW, B. S., Washburn College, 1923; M.S., University of Kansas, 1924; Instructor in Physics, Kansas State Agricultural College, 1924–27. American Telephone and Telegraph Company, Department of Development and Research, 1927–34; Bell Telephone Laboratories, 1934–. Mr. Barstow has been engaged in the development of methods and means of measuring and evaluating noise.

P. W. BLYE, S.B. in Electrical Engineering, Massachusetts Institute of Technology, 1919. American Telephone and Telegraph Company, Engineering Department, 1919; Department of Development and Research, 1919–34. Bell Telephone Laboratories, 1934–. Mr. Blye has been engaged in studies of the inductive coordination of power and telephone systems from the noise standpoint.

R. M. BOZORTH, A.B., Reed College, 1917; U. S. Army, 1917–19; Ph.D. in Physical Chemistry, California Institute of Technology, 1922; Research Fellow in the Institute, 1922–23. Bell Telephone Laboratories, 1923–. As Research Physicist, Dr. Bozorth is engaged in research work in magnetics.

R. M. BURNS, A.B., University of Colorado, 1915; A.M., 1916; Ph.D., Princeton University, 1921; Instructor, University of Colorado, 1916–17. Second Lieutenant, Chemical Warfare Service, U. S. Army, 1918–19. Research chemist, Barrett Company, 1921–22. Western Electric Company, 1922–25. Bell Telephone Laboratories, 1925–; Assistant Chemical Director, 1931–. Dr. Burns' work has been largely in the electrochemical field and particularly on the subject of the corrosion of metals and its prevention.

READ H. CARD, B.S. in Electrical Engineering, University of Tennessee, 1919. American Telephone and Telegraph Company, Long Lines

Department, 1919 and 1921–. Mr. Card's work has been concerned with transmission and inductive coordination matters.

G. W. ELMEN, B.Sc., University of Nebraska, 1902; M.A., 1904; D. Engg. (hon.), 1932. Research Laboratories of the General Electric Company, 1904–06; Engineering Department of the Western Electric Company, 1906–25; Bell Telephone Laboratories, 1925–. As Research Physicist, Dr. Elmen is engaged in magnetic research.

ALBERT G. GANZ, M.E., Stevens Institute of Technology, 1924; M.A., Columbia University, 1931. Engineering Department, Western Electric Company, 1924–25; Bell Telephone Laboratories, 1925–. Mr. Ganz has been engaged in the development of communication coils and transformers.

W. M. GOODALL, B.S. in Science, California Institute of Technology, 1928. Bell Telephone Laboratories, 1928–. Mr. Goodall has worked on ionosphere studies as well as general radio problems.

A. E. HARPER, M.E., Stevens Institute of Technology, 1922. Radio Intercept Service, U. S. Army, 1918–19. Western Electric Company, Engineering Department, 1922–23; American Telephone and Telegraph Company, Department of Development and Research, 1923–34; Bell Telephone Laboratories, 1934–. Mr. Harper has been engaged in radio transmission investigations with special reference to long-wave communication.

H. E. KENT, S.B. in Electrical Engineering, Massachusetts Institute of Technology, 1923; S.M., Massachusetts Institute of Technology, 1924. National Electric Light Association, 1924–33; Edison Electric Institute, 1933–. Mr. Kent has been engaged chiefly in work relating to the inductive coordination of power and communication circuits.

ARTHUR G. LAIRD, A.B., Harvard, 1916; S.B. in Electrical Engineering, Harvard, 1921. Engineering Department of the Western Electric Company, 1921–25; Bell Telephone Laboratories, 1925–. Mr. Laird has been engaged principally in the development of communication coils and transformers.

VICTOR E. LEGG, B.A., 1920, M.S., 1922, University of Michigan. Research Department, Detroit Edison Company, 1920–21; Bell Telephone Laboratories 1922–. Mr. Legg has been engaged in the development of magnetic materials and in their applications, par-

ticularly for the continuous loading of cables, and for compressed dust cores.

J. P. SCHAFER, B.S. in Electrical Engineering, Cooper Union, 1921; E.E., Cooper Union, 1925. Western Electric Company, 1915–25; Bell Telephone Laboratories, 1925–. Since 1928 Mr. Schafer has devoted a large part of his time to radio studies of the ionosphere.

S. A. SCHELKUNOFF, B.A., M.A., in Mathematics, The State College of Washington, 1923; Ph.D. in Mathematics, Columbia University, 1928. Engineering Department, Western Electric Company, 1923–25. Bell Telephone Laboratories, 1925–26. Department of Mathematics, State College of Washington, 1926–29. Bell Telephone Laboratories, 1929–. Dr. Schelkunoff has been engaged in mathematical research, especially in the field of electromagnetic theory.

A. M. SKELLETT, A.B., 1924, M.S., 1927, Washington University; Ph.D., Princeton University, 1933; Instructor, 1927–28, Assistant Professor of Physics, 1928–29, University of Florida. Bell Telephone Laboratories 1929–. Dr. Skellett, formerly engaged in investigations pertaining to the transatlantic radio telephone, is concerned with applications of electronic and ionic phenomena.